



Regeling van het College voor toetsen en examens van 28 november 2022, nummer CvTE-22.00976, houdende vaststelling van het beoordelingskader voor de doorstroomtoets in het primair onderwijs (Regeling beoordelingskader doorstroomtoets PO)

Het College voor toetsen en examens,

Gelet op artikel 3a, eerste lid, onderdeel g, van de Wet College voor toetsen en examens;

Gezien de goedkeuring van de Minister voor Primair en Voortgezet Onderwijs, gegeven op 4 november 2022, nummer 1303595,

Besluit:

Artikel 1. Beoordelingskader

Het beoordelingskader voor de doorstroomtoets als bedoeld in artikel 3a, eerste lid, onderdeel g van de Wet College voor toetsen en examens wordt vastgesteld als opgenomen in de bijlage van deze regeling.

Artikel 2. Technische specificaties levering van gegevens voor de beoordeling van de doorstroomtoets

De technische specificaties voor de levering van de gegevens voor de beoordeling van de doorstroomtoetsen, als bedoeld in artikel 1, worden beschreven in het handboek doorstroomtoets en het handboek normering of aanvullingen daarop zoals gepubliceerd op de website van het College voor toetsen en examens (cvte.nl/onderwerpen/toetsen-primair-onderwijs).

Artikel 3. Inwerkingtreding

Deze regeling treedt in werking op het tijdstip waarop artikel VIII, onderdeel B, van de Wet van 9 februari 2022 tot wijziging van een aantal onderwijswetten in verband met aanpassingen op het gebied van de doorstroom van het basisonderwijs naar het voortgezet onderwijs en wijziging van de stelselinrichting van doorstroomtoetsen en toetsen verbonden aan leerling- en onderwijsvolgsystemen in het basisonderwijs (Stb. 2022, 135) in werking treedt.

Artikel 4. Citeertitel

Deze regeling wordt aangehaald als: Regeling beoordelingskader doorstroomtoets PO.

Deze regeling zal met de toelichting in de Staatscourant worden geplaatst.

*Het College voor toetsen en examens,
de voorzitter,
J.H. van der Vegt*



BIJLAGE 1 BEOORDELINGSKADER VOOR DE DOORSTROOMTOETS

Bijlage behorende bij artikel 1 van de Regeling beoordelingskader doorstroomtoets PO

1. Inleiding

1.1. Begrippen en definities

Aanbieder: aanbieder van een doorstroomtoets.

Absolute normen: standaarden waaraan de vaardigheidsniveaus van leerlingen gerelateerd kunnen worden. Deze standaarden worden vastgesteld op basis van inhoudelijke en/of toetsspecifieke criteria.

Afnemer: een afnemer van een toets zoals een leerkracht, intern-begeleider, directeur of observator.

Algoritme (beslisregel): systeem van regels, logische of wiskundige handelingen, die als stappenplan gelden en kunnen worden uitgevoerd om tot een bepaald resultaat te komen.

Beoordelaarsovereenstemming: mate waarin verschillende beoordelaars tot identieke beoordelingen komen van (bijvoorbeeld) het werk van leerlingen. Twee veel gebruikte maten voor beoordelaarsovereenstemming zijn het percentage overeenstemming tussen twee beoordelaars en Cohen's Kappa.

Betrouwbaarheid: mate waarin de toetsscores vrij zijn van toevallige meetfouten.

CAT: Computergestuurde adaptieve toets op itemniveau.

Cesuur: minimale prestatie (gerepresenteerd als vaardigheid of toetsscore) die net indicatief is voor het halen van een (referentie)niveau en toetsadviescategorie.

Constructvaliditeit: eigenschap die een doorstroomtoets heeft als kan worden aangetoond dat de toets het door de constructeur beoogde kenmerk van de leerling (onderliggende trek, vaardigheid) meet. Deze wordt per kenmerk binnen een doorstroomtoets bepaald.

Dichotome scoring: wijze van scoren (punten toekennen) waarbij slechts twee waarderingen van het antwoord worden onderscheiden: bijvoorbeeld 'goed' of 'fout'. Een meerkeuzevragen wordt dichotoom gescoord (0 = niet correct, 1 = correct).

DIF: Differentieel item functioneren (DIF): een verschijnsel dat items verschillend functioneren voor bepaalde (sub)groepen leerlingen, zelfs wanneer deze (sub)groepen leerlingen een vergelijkbaar prestatieniveau hebben. Het gaat hierbij over de vraagzuiverheid, dat wil zeggen wanneer de verschillen in de prestaties op het item tussen groepen niet geheel verklaard kunnen worden door verschillen in de (latente) vaardigheid die aan de opgave ten grondslag ligt.

Discriminerend vermogen: mate waarin op grond van de scores op een item hoog- en laagscorende leerlingen (op de hele toets) onderscheiden kunnen worden.

Domein: verzameling van onderling samenhangende kennis en vaardigheden waarover getoetst wordt.

Effectgrootte: statistische maat om de grootte van een effect – bijvoorbeeld het verschil in de gemiddelde rekenprestaties tussen jongens en meisjes – weer te geven. Er wordt een onderscheid gemaakt tussen niet gestandaardiseerde (ruwe) effectgroottes en gestandaardiseerde effectgroottes. De bekendste gestandaardiseerde effectmaat is Cohen's d. Deze maat geeft het verschil tussen twee gemiddelden in het aantal standaarddeviaties. Ook de correlatie kan worden gezien als een gestandaardiseerde maat voor de effectgrootte.

Gezamenlijk anker: verzameling van items die in alle doorstroomtoetsen wordt opgenomen om verschillen in de moeilijkheidsgraad tussen de betreffende doorstroomtoetsen vast te kunnen stellen.

Inhoudvaliditeit: eigenschap dat de operationalisering van de inhoud van de doorstroomtoets een representatieve weergave is van het te meten referentieniveau.

Itemkalibratie: schatten (berekenen) van de psychometrische eigenschappen (parameters) van een verzameling items met het doel om deze items op dezelfde schaal te brengen.

Leerdoel: beoogde resultaat van onderwijsleeractiviteiten. Als het beoogde leerresultaat wordt geformuleerd in termen van uiterlijke condities en waarneembaar leerlinggedrag spreekt men van 'concrete leerdoelen' of 'gedragsdoelen'. Als het beoogde leerresultaat wordt geformuleerd in termen van cognitief gedrag of van cognitieve structuren, spreekt men van cognitieve leerdoelen.

Lokale betrouwbaarheid: mate waarin de toetsscores op een specifiek deel van de meetschaal vrij zijn van toevallige meetfouten.

Moeilijkheidsgraad: in hoeverre leerlingen uit de doelpopulatie in staat zijn om de vraag of vragen goed te beantwoorden. De moeilijkheid wordt meestal uitgedrukt met een p-waarde.

MST: adaptieve doorstroomtoets op moduleniveau, oftewel een multistage toets (MST). Bij een MST bepaalt een beslisregel het niveau van de leerling op basis van diens antwoorden op een set opgaven. Daarna krijgt de leerling een nieuwe set opgaven, afgestemd op het eerder bepaalde niveau.

Normeren: toekennen van een waardering aan een score. Bij de landelijke (toetsoverstijgende) normering van de doorstroomtoets betreft dit het vaststellen van de cesuren van de referentieniveaus en de toetsadviescategorieën.

PDCA-cyclus: Plan, Do, Check, Act cyclus is een model om de continue verbetering en vernieuwing in een organisatie te sturen. Het cyclisch karakter zorgt ervoor dat de kwaliteitsverbetering continu onder de aandacht is.

Po: primair onderwijs



Polytome scoring: scoren van een antwoord waarbij er meer dan twee onderscheidingen zijn in de waardering van het antwoord. Aan het antwoord op een vraag kunnen bijvoorbeeld 0, 1 of 2 punten worden toegekend. Bij open vragen hanteert men vaak polytome scoring.

Populatie: verzameling van vooraf gedefinieerde eenheden, bijvoorbeeld alle aan een doorstroomtoets deelnemende leerlingen uit groep 8 po, waarover men met behulp van statistisch onderzoek uitspraken wil doen.

Pretesten: afnemen van een aantal opgaven bij een groep leerlingen, meestal met het doel de eigenschappen van de doorstroomtoets te onderzoeken. Deze groep leerlingen is zoveel mogelijk vergelijkbaar met de populatie waarvoor de doorstroomtoets oorspronkelijk is ontwikkeld.

Referentiekader: beschrijving van wat leerlingen op verschillende momenten in hun schoolloopbaan op het gebied van Nederlandse taal en rekenen moeten kennen en kunnen (zie ook: Besluit referentieniveaus Nederlandse taal en rekenen).

Referentieniveaus: beschrijvingen van beheersingsdoelen voor Nederlandse taal en rekenen die aangeven wat leerlingen moeten kennen en kunnen op verschillende momenten tijdens een schoolloopbaan. De referentieniveaus zijn wettelijk verankerd in de Wet referentieniveaus Nederlandse taal en rekenen.

Schaal: reeks getallen die volgens een bepaald voorschrift gekoppeld worden aan waarnemingen.

Steekproef: verzameling leerlingen met specifieke kenmerken die op aselecte wijze uit de te onderzoeken populatie zijn getrokken.

Stopcriterium: aanpassen van de toetslengte. Een doorstroomtoets kan voor wat betreft het aantal toetsopgaven adaptief zijn in toetslengte. Het aanpassen van de toetslengte wordt in het geval van een computergestuurde adaptieve toets op itemniveau (CAT) geautomatiseerd afgehandeld. Na iedere opgave, wordt dan gekeken of de toets al gestopt kan worden. Stoppen gebeurt, zodra de vaardigheid van een leerling met de vereiste precisie te schatten is.

Terrein: dat in de doorstroomtoets aan bod dient te komen. Het referentiekader onderscheidt twee centrale terreinen: het terrein Nederlandse taal en het terrein rekenen. Beide terreinen worden uitgewerkt in onderliggende domeinen.

Toetsadvies: op basis van de behaalde resultaten op de doorstroomtoets berekende best passende niveau voor het volgen van aansluitend voorgezet onderwijs voor de leerling.

Toetsmatrijs: schematische representatie van het meetdoel.

Toetsreglement: document waarin de wijze van afnemen wordt geregeld.

Toetsterm: concrete, eenduidige en meetbare operationalisering van het betreffende leerdoel zoals beschreven in het bijpassende referentieniveau Nederlandse taal en rekenen.

1.2. Reikwijdte

Sinds het schooljaar 2014–2015 zijn scholen in het primair onderwijs verplicht in groep 8 een eindtoets Nederlandse taal en rekenen af te nemen. Per 1 januari 2023 wordt de eindtoets vervangen door de doorstroomtoets. De Wet doorstroomtoets po schrijft voor dat het College voor Toetsen en Examens (CvTE) als wettelijke taak krijgt om doorstroomtoetsen te erkennen. Doorstroomtoetsen die het CvTE erkent, worden voor een periode van vier jaar toegelaten tot het primair onderwijs.

Het CvTE maakt voor het erkennen van een Doorstroomtoets gebruik van een adviseur. De adviseur gebruikt voor zijn advies de Regeling beoordelingskader doorstroomtoetsen po.

Op basis van het advies van een adviseur stelt het CvTE jaarlijks vast of de erkende doorstroomtoets nog voldoet aan de criteria van de Regeling beoordelingskader doorstroomtoetsen po. Geconstateerde tekortkomingen kunnen leiden tot intrekken van de erkenning. Tegen deze intrekking kan de aanbieder bezwaar aantekenen.

Voor een lineaire doorstroomtoets geldt dat alle toetsopgaven, behalve die van het gezamenlijk anker en die van het eigen intern anker, jaarlijks moeten worden verversd.

Voor een computergestuurde adaptieve doorstroomtoets op itemniveau (zoals bij de CAT) geldt dat de itembank minimaal zes keer de omvang van een maximale operationele toets-versie dient te hebben. De aanbieder vermeldt daartoe, als onderdeel van de psychometrische aspecten, de maximale toetslengte in aantal toetsopgaven aangevuld met het stopcriterium per domein. In beginsel volstaat voor een CAT een jaarlijks verversing van 20–30% van de opgaven.

Voor een adaptieve doorstroomtoets op moduleniveau (MST) hanteert de aanbieder jaarlijks een adequate verversingsstrategie voor de opgaven van de itembank, passend bij de mate van adaptiviteit.

Tijdens de jaarlijkse vaststelling wordt eveneens beoordeeld of de toetsaanstellingsprocedure door alle aanbieders correct en consistent is uitgevoerd. Hierbij wordt, naast de pretest, ook de psychometrische kwaliteit van de operationele afname van T-1 beoordeeld.



2. Format Regeling beoordelingskader doorstroomtoetsen po

Het beoordelingsformat bevat de volgende drie inhoudelijke onderdelen, die enkele onderliggende thema's bevatten:

- H.3. Onderwijskundige aspecten
- H.4. Organisatorische aspecten
- H.5. Psychometrische aspecten

3. Onderwijskundige aspecten

3.1. Inleiding

In de volgende paragrafen worden de vakinhoudelijke kwaliteitseisen uitgewerkt voor de wettelijk verplichte domeinen Lezen en Begrippenlijst en Taalverzorging, daaronder niet begrepen het subdomein Begrippenlijst, van het terrein Nederlandse taal en de wettelijk verplichte domeinen Getallen, Verhoudingen, Meten en Meetkunde en Verbanden van het terrein rekenen én de optionele domeinen Schrijven en Mondelinge taalvaardigheid en het eveneens optionele subdomein Begrippenlijst van het terrein Nederlandse taal, in lijn met het Toetsbesluit PO, de Wet doorstroomtoetsen po en zoals vastgesteld in het Besluit referentieniveaus Nederlandse taal en rekenen. Voor alle wettelijk verplichte en optionele domeinen zijn in de Regeling beoordelingskader doorstroomtoetsen po kwaliteitseisen geformuleerd.

Naast de wettelijk verplichte en optionele terreinen en domeinen staat het de toetsaanbieder vrij om extra terreinen aan de doorstroomtoets toe te voegen, zoals: (sociaal-emotioneel) functioneren, luistervaardigheid en wereldoriëntatie. De score die de leerling op deze extra terreinen haalt, mag de toetsaanbieder toevoegen aan het leerlingrapport. Deze extra terreinen tellen niet mee voor het berekende toetsadvies en tellen ook niet mee voor de berekende score op de referentieniveaus. Alle toetsopgaven van de wettelijk verplichte, de optionele en de extra terreinen worden inhoudelijk door een adviseur van het CvTE beoordeeld. Zie in dit kader de kwaliteitseisen in paragraaf 3.2. De Regeling beoordelingskader doorstroomtoetsen po schrijft niet voor wat de minimale en maximale toetslengte in aantal toetsvragen van een doorstroomtoets dient te zijn. Eveneens worden er geen richtlijnen gegeven over de verhouding van het aantal toetsvragen tussen de wettelijk verplichte domeinen, optionele domeinen en extra domeinen. Verder worden er ook geen richtlijnen gegeven voor het aantal toetsvragen per referentieniveau 1F/2F/1S. Dit om de aanbieders de mogelijkheid te geven om zich van elkaar te onderscheiden en volgens de eigen zienswijze een compleet beeld van de leerlingen te kunnen geven.

3.2. Inhoudvaliditeit doorstroomtoets

Elementen van constructvaliditeit (bijvoorbeeld DIF onderzoek) en criterium validiteit (paragraaf 5.3) worden als onderdeel van de psychometrische aspecten beoordeeld.

Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
V1	Is de toetsmatrijs een adequate representatie van het meetdoel voor in ieder geval de wettelijk verplichte domeinen van Nederlandse taal en rekenen én de, indien van toepassing, optionele domeinen van Nederlandse taal?	ja/nee
V2	Voldoen de toetsopgaven van enerzijds de wettelijk verplichte domeinen van Nederlandse taal en rekenen én anderzijds de, indien van toepassing, optionele domeinen van Nederlandse taal in de toets aan de kwaliteitscriteria voor de inhoudvaliditeit (zijnde de relevantie van de inhoud van de toets) van toetsopgaven?	ja/nee
V3	Zijn de toetsopgaven van wettelijk verplichte domeinen van Nederlandse taal en rekenen én de, indien van toepassing, optionele domeinen van Nederlandse taal in de toets correct geconstrueerd?	ja/nee

Toelichting V1: De toetsmatrijs is voor wat betreft wettelijk verplichte domeinen van Nederlandse taal en rekenen én de, indien van toepassing, optionele domeinen van Nederlandse taal een adequate representatie van het meetdoel. Dat betekent eveneens dat er via de toetsmatrijs wordt voldaan aan de eisen uit de Toetswijzer PO.

Er is sprake van een adequate representatie wanneer de toetstermen het meetdoel presenteren. Dit blijkt uit het gegeven dat:

- de leerdoelen zijn geoperationaliseerd in observeerbare en meetbare toetstermen die ieder voor zich aansluiten bij de referentieniveaus voor de wettelijk verplichte en, indien van toepassing, optionele domeinen van de terreinen Nederlandse taal en rekenen;



- de toetsmatrijs in ieder geval bevat: (1) het aantal vragen met bijbehorende scorepunten per vraag, (2) de toetsvorm en/of het type vragen, (3) de verdeling van de verschillende vraag- en teksttypes in de toets, (4) de toegestane hulpmiddelen, en (5) de toetsduur;
- in het geval van een CAT er is verantwoord hoe de wettelijk verplichte en, indien van toepassing, optionele domeinen van de terreinen Nederlandse taal en rekenen op representatieve wijze voor alle individuele leerlingen aan bod komen.

Toelichting V2: Een adviseur evalueert in opdracht van het CvTE de relevantie van de inhoud van de toets van alle papieren toetsopgaven dan wel alle digitale items uit alle varianten van de doorstroomtoets en/of, indien van toepassing, uit de itembank van enerzijds de wettelijk verplichte domeinen van Nederlandse taal en rekenen én anderzijds van de, indien van toepassing, optionele domeinen van Nederlandse taal op de kwaliteitseisen: relevantie, objectiviteit, efficiëntie, specificiteit en neutraliteit, zoals beschreven in de Checklist voor het beoordelen van de kwaliteit van observatie-categorieën en toetsopgaven.

Indien in de doorstroomtoets ook optionele productieve vaardigheden worden getoetst, levert de aanbieder een beoordelaarsschema in, aangevuld met informatie over de beoordelaarsovereenstemming, een en ander conform de voorschriften uit de Toetswijzer PO.

Toelichting V3: Een adviseur evalueert in opdracht van het CvTE of de toetsopgaven van in ieder geval de wettelijk verplichte domeinen van Nederlandse taal en rekenen én de optionele domeinen van Nederlandse taal in de toets voldoen aan de constructievoorschriften voor toetsvragen, zoals beschreven in de Checklist voor het beoordelen van de kwaliteit van observatie-categorieën en toetsopgaven.

Beslisregel: Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, dienen alle vragen met JA te worden beantwoord.

3.3. Verantwoording

Kwaliteitseis

Code	Vraag	Mogelijke antwoorden
H1	Is er (een set van) verantwoordingsdocumenten aangeboden aan het CvTE met daarin voor wat betreft in ieder geval de wettelijk verplichte domeinen van Nederlandse taal en rekenen én, indien van toepassing, de optionele domeinen van Nederlandse taal een verantwoording van de gemaakte keuzes ten aanzien van de toetsinhoud, de afnamevorm, de psychometrische aspecten en de kwaliteitszorg rondom de itemconstructie van de toets?	ja/nee

Toelichting H1: Er wordt gecontroleerd of de verantwoordingsdocumenten voor wat betreft de wettelijk verplichte domeinen Nederlandse taal en rekenen én de, indien van toepassing, optionele domeinen van het terrein Nederlandse taal inhoudelijk voldoen aan de wettelijke kwaliteitseisen uit het Toetsbesluit PO en aan de inhoudelijke eisen van de Toetswijzer PO. Meer specifiek dient de door het CvTE te beoordelen set verantwoordingsdocumenten te bestaan uit de volgende (separate) documenten:

- Toetsreglement van de betreffende toets, inclusief informatie over de toetsinhoud en de afnamevorm;
- Wetenschappelijke verantwoording, inclusief de psychometrische aspecten van de doorstroomtoets, zoals beschreven in hoofdstuk 5 van deze regeling en zoals beschreven in de Regeling Beoordelingsnormen Doorstroomtoetsen po.

In het kader van transparantie, dient in de (set van) verantwoordingsdocumenten terug te vinden te zijn welke toetsonderdelen wel en niet meetellen voor het berekenen van het toetsadvies. Alle documenten dient de toetsaanbieder jaarlijks te updaten.

Beslisregel: Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, dient de (set van) verantwoordingsdocumenten (bewijslast) te worden beoordeeld als zijnde voldoende onderbouwd.

3.4. Terrein Nederlandse taal

In iedere doorstroomtoets worden voor wat betreft het terrein Nederlandse taal de wettelijk verplichte domeinen Lezen (§ 3.4.1) en Begrippenlijst en Taalverzorging, daaronder niet begrepen het subdomein Begrippenlijst, (§ 3.4.2) getoetst. Aanvullend hierop kunnen optioneel één of meerdere van de domeinen Schrijven (§ 3.4.3) en Mondelinge taalvaardigheid (§ 3.4.4) en het subdomein Begrippenlijst



(§ 3.4.5) worden getoetst. De gestelde vakinhoudelijke kwaliteitseisen voor deze vijf onderdelen en de vereiste verdeling van de toetsopgaven over de domeinen in de toetsmatrijs en de toetssamenstelling zijn opgenomen in de Toetswijzer PO.

3.4.1. Wettelijk verplicht domein: Lezen

Ongewenste gevoeligheid van teksten, zoals bv. reclame, wordt ondervangen met de kwaliteitseis Neutraliteit in de Checklist voor het beoordelen van de kwaliteit van observatie-categorieën en toetsopgaven. De toetsaanbieder is zelf verantwoordelijk voor het borgen van het auteursrecht op (delen) van de doorstroomtoets, dit aspect valt buiten deze regeling.

A. Subdomein Zakelijke Teksten

Kwaliteitseisen

Taken:

Code	Vragen	Mogelijke antwoorden
L1.1	Lezen de leerlingen informatieve teksten en maken ze daar opdrachten bij?	ja/nee
L1.2	Lezen de leerlingen instructieve teksten en maken ze daar opdrachten bij?	ja/nee
L1.3	Lezen de leerlingen betogende teksten en maken ze daar opdrachten bij?	ja/nee

Tekstkenmerken:

Code	Vragen	Mogelijke antwoorden
L2	Voldoen de teksten die de leerlingen lezen aan de tekstkenmerken inhoud (thema), vorm (taalgebruik en structuur) en bedoeling (doel), zoals gespecificeerd in Checklist 1 van deze regeling en zoals nader omschreven in het Referentiekader?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
L3.1	Wordt in de toets als kenmerk van de taakuitvoering techniek en woordenschat getoetst?	ja/nee
L3.2	Wordt in de toets als kenmerk van de taakuitvoering begrijpen getoetst?	ja/nee
L3.3	Wordt in de toets als kenmerk van de taakuitvoering interpreteren getoetst?	ja/nee
L3.4	Wordt in de toets als kenmerk van de taakuitvoering samenvatten getoetst?	ja/nee
L3.5	Wordt in de toets als kenmerk van de taakuitvoering opzoeken getoetst?	ja/nee
L3.6	Wordt in de toets als kenmerk van de taakuitvoering evalueren getoetst (conform de Toetswijzer PO is het onderwerp 'evalueren' een optioneel en niet verplicht onderdeel van de toets)?	ja/nee/n.v.t.
L3.7	Worden in de toets de verschillende kenmerken van de taakuitvoering evenwichtig getoetst binnen het subdomein zakelijke teksten?	ja/nee

Toelichting L1.1 t/m L1.3: Tenminste twee van de drie taken dienen in de doorstroomtoets te worden opgenomen.

Toelichting L3.7: Evenwichtig wil zeggen dat de in L3.1 tot en met L3.6 genoemde kenmerken gezamenlijk een representatieve weergave zijn van de inhoud van het betreffende subdomein van het terrein Nederlandse taal.

Toelichting L1 t/m L3: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010, p. 12–14) leidend.

B. Subdomein Fictionele, narratieve en literaire teksten

Kwaliteitseisen

Taken:



Code	Vragen	Mogelijke antwoorden
L4	Lezen de leerlingen fictionele, narratieve en/of literaire teksten en maken ze daar opdrachten bij?	ja/nee

Tekstkenmerken:

Code	Vragen	Mogelijke antwoorden
L5	Voldoen de teksten die de leerlingen lezen aan de tekstkenmerken inhoud (thema), vorm (taalgebruik en structuur) en bedoeling (doel), zoals gespecificeerd in Checklist 1 van deze regeling en zoals nader omschreven in het Referentiekader?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
L6.1	Wordt in de toets als kenmerk van de taakuitvoering begrijpen getoetst?	ja/nee
L6.2	Wordt in de toets als kenmerk van de taakuitvoering interpreteren getoetst?	ja/nee
L6.3	Optioneel: Wordt in de toets als kenmerk van de taakuitvoering evalueren getoetst (conform de Toetswijzer PO is het onderwerp 'evalueren' een optioneel en niet verplicht onderdeel van de toets)?	ja/nee/n.v.t.
L6.4	Worden in de toets de verschillende kenmerken van de taakuitvoering evenwichtig getoetst binnen het subdomein fictionele, narratieve en literaire teksten?	ja/nee

Toelichting L6.3: Evenwichtig wil zeggen dat de in L6.1 en L6.2 genoemde kenmerken gezamenlijk een representatieve weergave zijn van de inhoud van het betreffende subdomein van het terrein Nederlandse taal.

Beslisregel:

Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, gelden de volgende beslisregels:

1. Taken: van L1.1 t/m L1.3 dienen tenminste twee van de drie vragen met JA te worden beantwoord en dient L4 met JA te worden beantwoord.
2. Tekstkenmerken: vraag L2 en L5 dienen beide met JA te worden beantwoord.
3. Kenmerken van de taakuitvoering: alle vragen (L3.1 t/m L3.7 en L6.1 t/m L6.3) dienen met JA te worden beantwoord.

3.4.2. Wettelijk verplicht domein: Taalverzorging

Kwaliteitseisen

De leerlingen worden getoetst op verschillende categorieën van spellingsproblemen en -regels, zoals gespecificeerd in T1.1 t/m T1.7.

Code	Vragen	Mogelijke antwoorden
T1.1	Worden de leerlingen getoetst op regels voor lettergreepgrenzen?	ja/nee
T1.2	Worden de leerlingen getoetst op morfologische spelling?	ja/nee
T1.3	Worden de leerlingen getoetst op regels voor werkwoordspelling?	ja/nee
T1.4	Worden de leerlingen getoetst op overige regels?	ja/nee
T1.5	Worden de leerlingen getoetst op leestekens?	ja/nee
T1.6	Verantwoordt de toetsaanbieder dat er een evenwichtige verdeling is tussen opgaven rond niet-werkwoordspelling, werkwoordspelling en interpunctie?	ja/nee
T1.6a	In geval van een digitale adaptieve toets (CAT of MST), verantwoordt de toetsaanbieder op basis van welke beslisregel (het algoritme in het geval van een CAT en het module-design in geval van een MST) er wordt gegarandeerd dat er een evenwichtige verdeling is tussen opgaven rond niet-werkwoordspelling, werkwoordspelling en interpunctie?	ja/nee/n.v.t.

Toelichting T1.6: Evenwichtig wil zeggen dat de in T1.1 tot en met T1.5 genoemde categorieën van spellingsregels gezamenlijk een representatieve weergave zijn van de inhoud van het betreffende subdomein van het terrein Nederlandse taal.

Toelichting T1.6a: Naast het algoritme of module-design is de werking van de beslisregel aan te tonen



door middel van het opleveren van enkele toetspaden, zoals die door de beslisregel worden gegene-reerd.

Beslisregel:

Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, dienen de vragen T1.1 en T1.2 t/m T1.6 met JA te worden beantwoord.

3.4.3. Optioneel domein: Schrijven

Kwaliteitseisen schrijven via indirecte meting

Taken:

Code	Vragen	Mogelijke antwoorden
S1.1	Lezen de leerlingen tekstsoorten uit de schrijftakencategorie: Correspondentie (om te reviseren of te beoordelen)?	ja/nee
S1.2	Lezen de leerlingen tekstsoorten uit de schrijftakencategorie: Formulieren, berichten, advertenties, aantekeningen (om te reviseren of te beoordelen)?	ja/nee
S1.3	Lezen de leerlingen tekstsoorten uit de schrijftakencategorie: Verslagen, werkstukken, samenvattingen, artikelen (om te reviseren of te beoordelen)?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
S2.1	Maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect: samenhang?	ja/nee
S2.2	Maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect: afstemming op doel?	ja/nee
S2.3	Maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect: afstemming op publiek?	ja/nee
S2.4	Maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting S1.1 t/m S1.3: Tenminste twee van de drie bij de taken genoemde categorieën dienen in de doorstroomtoets te worden opgenomen.

Toelichting S1 en S2: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

Beslisregel schrijven indirecte meting:

Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, gelden de volgende beslisregels:

1. Taken: van S1.1 t/m S1.3 dienen tenminste twee van de drie vragen met JA te worden beantwoord.
2. Kenmerken van de taakuitvoering: de vragen bij S2 dienen met JA te worden beantwoord.

Kwaliteitseisen schrijven directe meting

Taken:

Code	Vragen	Mogelijke antwoorden
S3.1	Schrijven de leerlingen een tekst uit de schrijftakencategorie: Correspondentie?	ja/nee
S3.2	Schrijven de leerlingen een tekst uit de schrijftakencategorie: Formulieren, berichten, advertenties en/of aantekeningen?	ja/nee
S3.3	Schrijven de leerlingen een tekst uit de schrijftakencategorie: Verslagen, werkstukken, samenvattingen en/of artikelen?	ja/nee

Kenmerken van de taakuitvoering:



Code	Vragen	Mogelijke antwoorden
S4.1	Worden de teksten die de leerlingen schrijven beoordeeld op het aspect: samenhang?	ja/nee
S4.2	Worden de teksten die de leerlingen schrijven beoordeeld op het aspect: afstemming op doel?	ja/nee
S4.3	Worden de teksten die de leerlingen schrijven beoordeeld op het aspect: afstemming op publiek?	ja/nee
S4.4	Worden de teksten die de leerlingen schrijven beoordeeld op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting S3.1 t/m S3.3: Minimaal één taak per categorie in de doorstroomtoets dient te worden opgenomen.

Toelichting S3 en S4: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

Beslisregel schrijven directe meting:

Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, gelden de volgende beslisregels:

1. Taken: De vragen bij S3.1 t/m S3.3 dienen alle met JA te worden beantwoord (minimaal één taak per categorie).
2. Kenmerken van de taakuitvoering: de vragen bij S4 dienen met JA te worden beantwoord.

3.4.4. Optioneel domein: Mondelinge taalvaardigheid

Kwaliteitseisen mondelinge taalvaardigheid via indirecte meting

A. Subdomein Gesprekken

Kwaliteitseisen

Taken:

Code	Vragen	Mogelijke antwoorden
M1.1	Luisteren en beoordelen de leerlingen gesprekken uit de takencategorie: discussie en overleg?	ja/nee
M1.2	Luisteren en beoordelen de leerlingen gesprekken uit de takencategorie: informatie uitwisselen?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M2.1	Maken leerlingen opgaven waarin ze gesprekken beoordelen op het aspect: beurten nemen en bijdragen aan samenhang?	ja/nee
M2.2	Maken leerlingen opgaven waarin ze gesprekken beoordelen op het aspect: afstemming op doel?	ja/nee
M2.3	Maken leerlingen opgaven waarin ze gesprekken beoordelen op het aspect: afstemming op de gesprekspartners?	ja/nee
M2.4	Maken leerlingen opgaven waarin ze gesprekken beoordelen op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting M1.1 en M1.2: Tenminste één van de twee taken dient in de doorstroomtoets te worden opgenomen.

Toelichting M1 en M2: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

Beslisregel mondelinge taalvaardigheid indirecte meting, subdomein Gesprekken:

Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, gelden de volgende beslisregels:

1. Taken: van M1.1 en M1.2 dient één van de twee vragen met JA te worden beantwoord (taak is minimaal één keer aanwezig).
2. Kenmerken van de taakuitvoering: de vragen bij M2 dienen met JA te worden beantwoord.



B. Subdomein Spreken

Kwaliteitseisen

Taken:

Code	Vragen	Mogelijke antwoorden
M3	Luisteren en beoordelen de leerlingen een monoloog?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M4.1	Maken leerlingen opgaven waarin ze een monoloog beoordelen op het aspect: samenhang?	ja/nee
M4.2	Maken leerlingen opgaven waarin ze een monoloog beoordelen op het aspect: afstemming op doel?	ja/nee
M4.3	Maken leerlingen opgaven waarin ze een monoloog beoordelen op het aspect: afstemming op publiek?	ja/nee
M4.4	Maken leerlingen opgaven waarin ze een monoloog beoordelen op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting M3 en M4: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

Beslisregel mondelinge taalvaardigheid indirecte meting, subdomein Spreken:

Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, gelden de volgende beslisregels:

1. Taken: vraag M3 dient met JA te worden beantwoord.
2. Kenmerken van de taakuitvoering: de vragen bij M4 dienen met JA te worden beantwoord.

Kwaliteitseisen mondelinge taalvaardigheid via directe meting

A. Subdomein Gesprekken

Kwaliteitseisen

Taken:

Code	Vragen	Mogelijke antwoorden
M5.1	Nemen de leerlingen deel aan gesprekken uit de takencategorie: discussie en overleg?	ja/nee
M5.2	Nemen de leerlingen deel aan gesprekken uit de takencategorie: informatie uitwisselen?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M6.1	Worden de gesprekken die leerlingen voeren beoordeeld op het aspect: beurten nemen en bijdragen aan samenhang?	ja/nee
M6.2	Worden de gesprekken die leerlingen voeren beoordeeld op het aspect: afstemming op doel?	ja/nee
M6.3	Worden de gesprekken die leerlingen voeren beoordeeld op het aspect: afstemming op de gesprekspartners?	ja/nee
M6.4	Worden de gesprekken die leerlingen voeren beoordeeld op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting M5.1 en M5.2: Tenminste één van de twee taken dient in de doorstroomtoets te zijn opgenomen.

Toelichting M5 en M6: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

Beslisregel mondelinge taalvaardigheid directe meting, subdomein Gesprekken:



Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, gelden de volgende beslisregels:

1. Taken: van M5.1 en M5.2 dient één van de twee vragen met JA te worden beantwoord (taak is minimaal één keer aanwezig).
2. Kenmerken van de taakuitvoering: de vragen bij M6 dienen met JA te worden beantwoord.

B. Subdomein Luisteren

Kwaliteitseisen

Taken:

Code	Vragen	Mogelijke antwoorden
M7.1	Luisteren de leerlingen naar instructies?	ja/nee
M7.2	Luisteren de leerlingen als lid van een live publiek?	ja/nee
M7.3	Luisteren de leerlingen naar radio en/of televisie en/of naar gesproken tekst op internet en/of naar overige door de aanbieder gehanteerde communicatiemiddelen?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M8.1	Wordt tijdens het luisteren getoetst of leerlingen de luistertekst kunnen begrijpen?	ja/nee
M8.2	Wordt tijdens het luisteren getoetst of leerlingen de luistertekst kunnen interpreteren?	ja/nee
M8.3	Wordt tijdens het luisteren getoetst of leerlingen de luistertekst kunnen samenvatten?	ja/nee

Toelichting M7.1 t/m M7.3: Twee van de drie taken dienen in de doorstroomtoets te worden opgenomen.

Toelichting bij M7 en M8: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

Beslisregel mondelinge taalvaardigheid indirecte meting, subdomein Luisteren:

Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, gelden de volgende beslisregels:

1. Taken: van M7.1 t/m M7.3 dienen twee van de drie vragen met JA te worden beantwoord (taak is minimaal één keer aanwezig).
2. Kenmerken van de taakuitvoering: de vragen bij M8 dienen met JA te worden beantwoord.

C. Subdomein Spreken

Kwaliteitseisen

Taken:

Code	Vragen	Mogelijke antwoorden
M9	Houden de leerlingen een monoloog?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M10.1	Wordt de monoloog die de leerlingen houden beoordeeld op het aspect: samenhang?	ja/nee
M10.2	Wordt de monoloog die de leerlingen houden beoordeeld op het aspect: afstemming op doel?	ja/nee
M10.3	Wordt de monoloog die de leerlingen houden beoordeeld op het aspect: afstemming op publiek?	ja/nee
M10.4	Wordt de monoloog die de leerlingen houden beoordeeld op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting M9 en M10: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.



Beslisregel mondelinge taalvaardigheid indirecte meting, subdomein Spreken:

Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, gelden de volgende beslisregels:

1. Taken: vraag M9 dient met JA te worden beantwoord.
2. Kenmerken van de taakuitvoering: de vragen bij M10 dienen met JA te worden beantwoord.

3.4.5. Optioneel domein: Begrippenlijst

Kwaliteitseis

Code	Vragen	Mogelijke antwoorden
B1	Worden de leerlingen getoetst met opgaven die betrekking hebben op de begrippenlijst?	ja/nee

Toelichting B1: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010, p. 17–19) leidend.

Beslisregel Begrippenlijst:

Om voor deze kwaliteitseis een voldoende te kunnen krijgen, dient vraag B1 met JA te worden beantwoord.

3.5. Terrein rekenen

Het Referentiekader onderscheidt voor het terrein rekenen vier wettelijk verplichte domeinen, te weten Getallen (g), Verhoudingen (vh), Meten en meetkunde (m/mk) en Verbanden (vb). In de doorstroomtoets dienen alle domeinen getoetst te worden. Dit betekent dat alle doorstroomtoetsen moeten voldoen aan de gestelde kwaliteitseisen aan het terrein rekenen. De gestelde kwaliteitseisen voor de vier domeinen zijn opgenomen in de Toetswijzer PO.

Een individuele rekenopgave kan betrekking hebben op meerdere domeinen van het terrein rekenen. In dat geval dient de aanbieder dit duidelijk te vermelden, bijvoorbeeld door te kiezen voor het domein dat het beste past bij het beoogde toetsdoel van de betreffende opgave. Tevens dient de aanbieder dan toe te lichten op welke wijze er wordt voldaan aan de voorgeschreven verdeling van toetsopgaven over de domeinen van het terrein rekenen.

3.5.1. Verdeling toetsopgaven over domeinen in toetsmatrijs en toetsamenstelling

De doorstroomtoets bevat opgaven uit alle domeinen uit het referentiekader rekenen. Het betreft de domeinen Getallen (g), Verhoudingen (vh), Meten en meetkunde (m/mk) en Verbanden (vb).

Kwaliteitseisen

De opgaven dienen in de toets als volgt over de domeinen te zijn verdeeld:

Code	Vragen	Mogelijke antwoorden
D.G	Ligt het percentage toetsopgaven uit het domein Getallen tussen de 30% en 40% van het totaal aantal toetsopgaven van het terrein rekenen?	ja/nee
D.Vh	Ligt het percentage toetsopgaven uit het domein Verhoudingen tussen de 20% en 30% van het totaal aantal toetsopgaven van het terrein rekenen?	ja/nee
D.M/mk	Ligt het percentage toetsopgaven uit het domein uit het domein Meten & meetkunde tussen de 20% en 30% van het totaal aantal toetsopgaven van het terrein rekenen?	ja/nee
D.Vb	Ligt het percentage toetsopgaven uit het domein Verbanden tussen de 15% en 20% van het totaal aantal toetsopgaven van het terrein rekenen?	ja/nee

Toelichting: De gestelde percentages toetsopgaven per domein gelden voor zowel papieren toetsen, lineaire digitale toetsen, adaptieve toetsen op itemniveau (CAT) en adaptieve toetsen op modulenniveau (MST). Toetsaanbieders van adaptieve toetsen (CAT en MST) verantwoorden op basis van voorbeelden van toetspaden, zoals die door de beslisregel (het algoritme) worden gegenereerd, de vereiste verdeling van toetsopgaven over de domeinen.

Beslisregel: Om deze categorie van kwaliteitseisen met een voldoende te kunnen afsluiten, dienen alle vragen met JA te worden beantwoord.



3.5.2. Verdeling toetsopgaven over onderdelen

Elk domein is opgebouwd uit drie onderdelen (zie ook Checklist 3):

- Notatie, taal en betekenis, waarbij het gaat om de uitspraak, schrijfwijze en betekenis van getallen, symbolen en relaties en om het gebruik van wiskundetaal.
- Met elkaar in verband brengen, waarbij het gaat om het verband tussen begrippen, notaties, getallen en dagelijks spraakgebruik.
- Gebruiken, waarbij het er om gaat rekenkundige vaardigheden in te zetten bij het oplossen van problemen.

De toets bevat opgaven uit alle onderdelen.

Kwaliteitseisen

De relatieve (procentuele) verdeling¹ van opgaven over de onderscheiden onderdelen is als volgt:

Code	Vragen	Mogelijke antwoorden
O.A	Is iedere toetsopgave, met uitzondering van onderdeel A, onder één onderdeel geplaatst ¹ ?	ja/nee
O.B	Is het percentage eigen toetsopgaven uit onderdeel B afgerond ten minste 20 procent van het totaal aantal toetsopgaven van het terrein rekenen?	ja/nee
O.C	Is het percentage toetsopgaven uit onderdeel C afgerond ten minste 30 procent van het totaal aantal toetsopgaven van het terrein rekenen?	ja/nee
O.BC	Is het percentage toetsopgaven uit onderdeel C hoger dan het percentage toetsopgaven uit onderdeel B?	ja/nee

¹ Voor onderdeel A is geen minimum vereist, omdat deze inhouden ook kunnen worden gebruikt en toegepast bij de inhouden van B en C.

Toelichting: De gestelde afgeronde percentages toetsopgaven per domein gelden voor zowel papieren toetsen, lineaire digitale toetsen, computergestuurde adaptieve toetsen op itemniveau (CAT) en adaptieve toetsen op moduleniveau (MST). Naast het algoritme of module-design is door aanbieders van adaptieve toetsen (CAT en MST) de werking van de beslisregel aan te tonen door middel van het opleveren van enkele toetspaden die voldoen aan de vereiste verdeling van toetsopgaven over de domeinen, zoals die door de beslisregel worden gegenereerd.

Beslisregel: Om deze categorie van kwaliteitseisen met een voldoende te kunnen afsluiten, dienen alle vragen met JA te worden beantwoord.

3.5.3. Opgaven met en zonder context

In het terrein rekenen van een doorstroomtoets moeten zowel opgaven met context als opgaven zonder context (zogenoemde 'kale opgaven') worden opgenomen, een en ander conform de Toetswijzer PO.

Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
C1	Is het percentage toetsopgaven van het type contextopgaven voor het domein Getallen ten minste 30 procent van het totaal aantal toetsopgaven van het domein Getallen?	ja/nee
C2	Is het percentage toetsopgaven van het type contextloze opgaven voor het domein Getallen ten minste 20 procent van het totaal aantal toetsopgaven van het domein Getallen?	ja/nee
C3	Is het percentage toetsopgaven van het type contextopgaven voor het domein Verhoudingen ten minste 30 procent van het totaal aantal toetsopgaven van het domein Verhoudingen?	ja/nee
C4	Is het percentage toetsopgaven van het type contextloze opgaven voor het domein Verhoudingen ten minste 20 procent van het totaal aantal toetsopgaven van het domein Verhoudingen?	ja/nee

Toelichting: De gestelde percentages toetsopgaven per domein gelden voor zowel papieren toetsen, lineaire digitale toetsen, computergestuurde adaptieve toetsen op itemniveau (CAT) en adaptieve toetsen op moduleniveau (MST). Toetsaanbieders van adaptieve toetsen (CAT en MST) tonen naast

¹ Gebruik hierbij checklist 2.



het algoritme of module-design met de werking van de beslisregel aan door middel van het opleveren van enkele toetspaden, zoals die door de beslisregel worden gegenereerd de vereiste verdeling van toetsopgaven over de domeinen aan.

Beslisregel: Om deze categorie van kwaliteitseisen met een voldoende te kunnen afsluiten, dienen alle vragen met JA te worden beantwoord.

3.5.4. Gebruik kladpapier

Kwaliteitseis

Code	Vragen	Mogelijke antwoorden
K1	Mogen de leerlingen bij het beantwoorden van de (digitale) toetsopgaven fysiek en/of niet programmeerbaar kladpapier gebruiken?	ja/nee
K2	Wordt in de handleiding voor leraren aangegeven dat zij leerlingen nadrukkelijk dienen te wijzen op het toegestane gebruik van fysiek en/of niet programmeerbaar kladpapier?	ja/nee

Beslisregel: Om deze categorie van kwaliteitseisen met een voldoende te kunnen afsluiten, dienen alle vragen met JA te worden beantwoord.

3.5.5. Gebruik rekenmachine

Kwaliteitseis

Code	Vragen	Mogelijke antwoorden
R1	Mogen de leerlingen bij het beantwoorden van de (digitale) toetsopgaven een niet-grafische en/of niet-programmeerbare rekenmachine gebruiken conform het in de Toetswijzer PO voorgeschreven aantal toetsopgaven van de doorstroomtoets?	ja/nee

Toelichting: Gebruik van een niet-grafische en/of niet-programmeerbare rekenmachine is, met een minimum van 0% tot het in de Toetswijzer PO voorgeschreven maximum aandeel van toetsopgaven, alleen toegestaan voor de grondbewerkingen optellen, aftrekken, vermenigvuldigen en/of delen. Gebruik van een grafische en/of programmeerbare rekenmachine is niet toegestaan. En de rekenmachine mag niet worden ingezet ter vervanging van het zelf rekenen.

Beslisregel: Om deze kwaliteitseis met een voldoende te kunnen afsluiten, dient deze vraag met JA te worden beantwoord.

3.6. Toetsopgaven Nederlandse taal en rekenen op niveau 1F en 2F/1S

In het document *Referentiekader taal en rekenen* (SLO, 2011) is vastgelegd wat leerlingen eind groep 8 van de basisschool moeten kennen en kunnen als het gaat om basisvaardigheden voor Nederlandse taal en rekenen. Deze kennis en vaardigheden worden in het referentiekader gespecificeerd in een aantal referentieniveaus. De verschillende niveaus worden beschreven in twee kwaliteiten: fundamentele kwaliteit (F) en streefkwaliteit (S). De niveaus zijn cumulatief. Een leerling moet op een hoger niveau alle vaardigheden beheersen die op een lager niveau genoemd worden. Deze worden niet telkens herhaald. De beschrijvingen van de niveaus voor Nederlandse taal en rekenen verschillen. Dit komt doordat de vakken verschillen.

Voor Nederlandse taal zijn vier niveaus beschreven: 1F, 2F, 3F en 4F en voor rekenen zijn zes niveaus beschreven: 1F, 2F, 3F en 1S, 2S en 3S. Voor het po zijn voor het terrein Nederlandse taal de niveaus 1F en 2F wettelijk vastgelegd. Voor het terrein rekenen zijn dit voor het po de niveaus 1F en 1S.

Dit betekent dat, in lijn met de Toetswijzer PO, de doorstroomtoets, ongeacht de toetsvorm, voor de terreinen Nederlandse taal en rekenen toetsopgaven moet bevatten die de inhoud van de niveaus 1F (fundamentele kwaliteit) en 2F / 1S (streefkwaliteit) meten.



Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
TN1	Zijn de opgaven van de toets voor de terreinen Nederlandse taal en rekenen breed samengesteld qua referentieniveaus 1F en 2F / 1S?	ja/nee

Toelichting TN1: Dit aspect dient uit de toetsmatrijs naar voren te komen. Dat wil zeggen dat in de toetsmatrijs de verhouding 1F en 2F / 1S vragen is gespecificeerd en dat in zowel de toetsmatrijs als in de varianten van de doorstroomtoets de verhouding 1F en 2F / 1S vragen evenwichtig is. Evenwichtig wil zeggen dat het aantal en de verhouding 1F en 2F / 1S vragen gezamenlijk inhoudelijk de referentieniveaus 1F en 2F/1S van de terreinen Nederlandse taal en rekenen representeren, een en ander conform de voorschriften uit de Toetswijzer PO.

Een CAT dient voor elke individuele leerling voor de terreinen Nederlandse taal en rekenen een uitspraak over het behaalde niveau 1F en/of 2F / 1S te geven. In het geval van een CAT is er verantwoord hoe het 1F en/of 2F/1S niveau van de terreinen Nederlandse taal en rekenen op representatieve wijze voor alle individuele leerlingen aan bod komt. De aanbieder kan dit verantwoorden aan de hand van de toetsmatrijs, aangevuld met de procesbeschrijving van hoe het algoritme werkt, aangevuld met enkele mogelijke itemsets inclusief de daarin aangeboden toetsopgaven en inclusief een toelichting op het daarbij geldende stopcriterium.

Elke routing binnen het MST design dient voor de terreinen Nederlandse taal en rekenen een uitspraak over het behaalde niveau 1F en/of 2F / 1S te geven. In het geval van een MST is er verantwoord hoe het 1F en/of 2F/1S niveau van de terreinen Nederlandse taal en rekenen op een bij de moeilijkheidsgraad van de betreffende route passende wijze voor alle individuele leerlingen aan bod komt. De aanbieder kan dit verantwoorden aan de hand van de toetsmatrijs, aangevuld met de beschrijving van de samenstelling van het MST design, aangevuld met enkele toetspaden uit het design inclusief de daarin aangeboden toetsopgaven.

Beslisregel: Om deze kwaliteitseis met een voldoende te kunnen afsluiten, dient vraag TN1 met JA te worden beantwoord.

4. Organisatorische aspecten

4.1. Leerlingrapport

Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
LR1	Stelt de toetsaanbieder voor elke individuele leerling die de doorstroomtoets aflegt een leerlingrapport op?	ja/nee
LR2	Bevat het leerlingrapport het toetsadvies, uitgedrukt in één van de toetsadviescategorieën?	ja/nee
LR3	Geeft het leerlingrapport aan welk referentieniveau Lezen de leerling beheerst?	ja/nee
LR4	Geeft het leerlingrapport aan welk referentieniveau Taalverzorging de leerling beheerst?	ja/nee
LR5	Geeft het leerlingrapport aan welk referentieniveau rekenen de leerling beheerst?	ja/nee
LR6	Bevat het leerlingrapport een toelichting bij het toetsadvies, de beheersing van de referentieniveaus en eventuele optionele getoetste onderdelen van de leerling die (ook) voor ouders, verzorgers, voogden en afnemers begrijpelijk is?	ja/nee

Toelichting LR1: De toetsaanbieder geeft één of meerdere (model)rapportages ter beoordeling. Wanneer er meer dan één variant van het leerlingrapport bestaat, moeten al deze varianten aangeboden worden, met een uitleg wanneer welke variant wordt gebruikt. Aangezien alle varianten aan de eisen moeten voldoen, worden alle voorbeeldrapporten beoordeeld op de kwaliteitseisen LR2 tot en met LR6. Mochten er rapporten ontbreken (d.w.z. LR1 = nee), dan krijgt de toetsaanbieder een onvoldoende voor 'leerlingrapport'. Wanneer het leerlingrapport wel wordt aangeboden, worden de kwaliteitseisen LR2 tot en met LR6 beoordeeld op basis van die versie(s) van het leerlingrapport.

Toelichting LR2: De toetsaanbieder dient in het leerlingrapport een toetsadvies te geven. De aanbieder hanteert hierbij de door de rijksoverheid vastgestelde toetsadviescategorieën. Dit toetsadvies is gebaseerd op in ieder geval de resultaten van de wettelijk verplichte domeinen Lezen, Taalverzorging en rekenen. Aanvullend mag de toetsaanbieder in de bepaling van het toetsadvies ook één of meerdere van de optionele domeinen van Nederlandse taal meenemen die in dit kader in Hoofdstuk 3



beschreven zijn. In de wetenschappelijke verantwoording moet zijn aangegeven op basis van metingen van welke domeinen het toetsadvies tot stand is gekomen.

Toelichting LR3: In het leerlingrapport moet aangegeven zijn op welk referentieniveau de leerling gepresteerd heeft op het domein Lezen. In het leerlingrapport moet in ieder geval aangegeven worden in welk van de volgende drie categorieën de leerlingprestatie past: onder niveau 1F, op niveau 1F, of op streefniveau 2F. Hoewel in de theorie de categorieën elkaar niet uitsluiten, en de leerling ook op een hoger niveau zou kunnen presteren dan het maximaal gegeven niveau, is in deze regeling vastgesteld dat in de rapportage een leerlingprestatie in één (niet meer en niet minder) van deze drie categorieën geclassificeerd wordt.

Toelichting LR4: In het leerlingrapport moet aangegeven zijn op welk referentieniveau de leerling gepresteerd heeft op het domein Taalverzorging. In het leerlingrapport moet in ieder geval aangegeven worden in welk van de volgende drie categorieën de leerlingprestatie past: onder niveau 1F, op niveau 1F, of op streefniveau 2F. Hoewel in de theorie de categorieën elkaar niet uitsluiten, en de leerling ook op een hoger niveau zou kunnen presteren dan het maximaal gegeven niveau, is in deze regeling vastgesteld dat in de rapportage een leerlingprestatie in één (niet meer en niet minder) van deze drie categorieën geclassificeerd wordt.

Toelichting LR5: In het leerlingrapport moet aangegeven zijn op welk referentieniveau de leerling gepresteerd heeft op het domein rekenen. In het leerlingrapport moet in ieder geval aangegeven worden in welk van de volgende drie categorieën de leerlingprestatie past: onder niveau 1F, op niveau 1F, of op streefniveau 1S. Hoewel in de theorie de categorieën elkaar niet uitsluiten, en de leerling ook op een hoger niveau zou kunnen presteren dan het maximaal gegeven niveau, is in deze regeling vastgesteld dat in de rapportage een leerlingprestatie in één (niet meer en niet minder) van deze drie categorieën geclassificeerd wordt.

Toelichting LR6: Er dient een geschreven toelichting te worden geboden, waarin begrijpelijke handvatten gegeven worden voor de interpretatie van het toetsadvies, de beheersing van de referentieniveaus en eventuele optionele getoetste onderdelen van de leerling door diverse betrokkenen. De aanbieder toont aan op welke wijze de rapportage rekening houdt met de verschillende doelgroepen.

Beslisregel: Om deze categorie van kwaliteitseisen met een voldoende te kunnen afsluiten, dienen alle vragen met JA te worden beantwoord.

4.2. Leerlingen

De doorstroomtoets is geschikt voor alle leerlingen, ook voor leerlingen met een specifieke ondersteuningsbehoefte, met uitzondering van leerlingen die onder de ontheffingsgronden² vallen.

Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
LL1	Kunnen alle leerlingen, ook met specifieke ondersteuningsbehoeften (uitgezonderd leerlingen die vallen onder de ontheffingsgronden), deelnemen aan de toets?	ja/nee
LL1.a	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met een visuele beperking?	ja/nee
LL1.b	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor dove of slechthorende leerlingen of leerlingen met een taalontwikkelingsstoornis?	ja/nee
LL1.c	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met dyslexie?	ja/nee
LL1.d	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met dyscalculie?	ja/nee
LL1.e	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met andere beperkingen dan genoemd in LL1.a t/m LL1.d?	ja/nee
LL1.f	Wordt beschreven wat de mogelijkheden zijn als de geboden ondersteuning niet toereikend is?	ja/nee
LL2	Kunnen alle leerlingen gebruikmaken van (de in Toetswijzer PO voorgeschreven) hulpmiddelen?	ja/nee

Toelichting LL1.a t/m LL1.e: Er is aantoonbaar aandacht besteed aan alle in LL1.a t/m LL1.e genoemde doelgroepen en hun ondersteuningsbehoeften. De vorm van de ondersteuning staat beschreven in de handleiding, waaronder de wijze waarop rekening gehouden wordt met doelgroepen in de itemcon-

² Stb 2022, 135 (artikel 45C)



structie, door het aanbod van varianten van de toets en/of via de afdnamecondities.

Bij de itemconstructie wordt generiek minstens op de volgende wijze rekening gehouden met toegankelijkheid: richtlijnen voor kleurgebruik en contrast bij items, richtlijnen met betrekking tot een ondubbelzinnige lay-out en tekstuele vormgeving, richtlijnen met betrekking tot figuurlijk taalgebruik (als dat geen toetsdoel is) en richtlijnen met betrekking tot het gebruik van ik, gebiedende wijs en vraagzinnen.

Toelichting LL 1.a: De toets wordt door de aanbieder (op bestelling) geleverd in een format dat zich leent voor de hulpmiddelen die door leerlingen met een visuele beperking worden gebruikt en met inachtneming van de toetseisen inhoudelijk geschikt zijn gemaakt voor de leerlingen met een visuele beperking. Afhankelijk van de aard van de visuele beperking kan tijdverlenging worden toegekend.

Bij de constructie van de toets is rekening gehouden met kleurenblinde kandidaten. Daarnaast is er bijvoorbeeld de mogelijkheid tot het aanbieden van een zwart/wit-variant. Relevant voor bij de normering is dat waar het ontbreken van een kleuren(afbeelding) invloed op de moeilijkheid van het item kan hebben, er dit aangegeven dient te worden, bijvoorbeeld door middel van een andere itemidentificatie dan de opgave waar deze op gebaseerd is.

Bij twijfel mag de leerling een beroep doen op de volgende procedure: voor kleurenblinde leerlingen kan een opzoekhulp worden ingezet die op verzoek van de leerling de kleur van een door de leerling aangewezen vlakdeel benoemt, of een door de leerling aangewezen kleur aanwijst in toets of hulpmiddel.

Het staat andere aanbieders vrij om voor (zeer) slechtziende en blinde leerlingen aanpassingen voor de toegankelijkheid van hun toets aan te bieden. Bijvoorbeeld door gebruik te maken van digitale brailleleesregels, tekenboeken of het aanbieden van een braillevariant. Indien een aanbieder hiervoor kiest verantwoordt de aanbieder de gekozen aanpassingen, waarbij zo min mogelijk wordt afgeweken van de reguliere toets.

Toelichting LL 1.b: De aanbieder levert bijvoorbeeld (op bestelling) een aangepaste toets indien de toets gebruikmaakt van functioneel geluid (zoals bijvoorbeeld bij dictee-opgaven).

Toelichting LL 1.c: De aanbieder geeft aan hoe bij de lay-out en typografie met deze doelgroep rekening wordt gehouden.

Als een leerling door dyslexie moeite heeft met het lezen van teksten op papier of beeldscherm, moet de leerling gebruik kunnen maken van een door de aanbieder geleverde voorleesfunctie. De wijze waarop kan per toets(soort) verschillen.

Andere voorbeelden van hulpmiddelen ter ondersteuning van deze doelgroep zijn de beschikbaarheid van een zoomfunctie en een markeerstift.

Toelichting bij LL 1.d: Voor leerlingen met dyscalculie wordt geen aangepaste toets geleverd. Ook zijn geen specifieke hulpmiddelen toegestaan. Wel kan in de afdnamecondities (bijvoorbeeld door het geven van meer tijd) rekening worden gehouden met deze ondersteuningsbehoefte. De mogelijkheden staan genoemd in de verantwoordingsdocumenten.

Toelichting bij LL 1.e: Voor leerlingen met andere beperkingen, bijvoorbeeld een motorische beperking, een autisme spectrum stoornis (ASS) of een aandachtstekortstoornis met of zonder hyperactiviteit (AD(H)D), zit de mogelijke aanpassing in de wijze van afnemen, bijvoorbeeld;

- tijdverlenging;
- inzetten van extra pauze(s);
- voorkomen van afleidingen tijdens de afdname;
- structureren en vooraf laten ervaren van de regels.

De aanbieder toont aan welke ondersteuning wordt geboden voor andere beperkingen zoals hier genoemd.

Toelichting LL 1.f: De aanbieder toont aan wat er aan maatwerk mogelijk is indien het beschreven aanbod aan hulpmiddelen, toetsvarianten en aanpassingen van afdnamecondities niet toereikend zijn voor een leerling.

In de Toetswijzer PO staat van een aantal hulpmiddelen voorgeschreven dat deze altijd zijn toegestaan of wanneer al dan niet deze zijn toegestaan. Aanvullend kan een aanbieder andere hulpmiddelen toestaan.



Toelichting LL2: De aanbieder schrijft voor dat bij alle toets(vorm)en alle onderdelen de hulpmiddelen zijn toegestaan binnen de gegeven kaders van de vigerende regeling, zoals vermeld in de Toetswijzer PO.

Beslisregel: Om deze categorie van kwaliteitseisen met een voldoende te kunnen afsluiten, dienen vraag LL1, LL1.a t/m LL1.f en LL2 met JA te worden beantwoord.

4.3. Afname doorstroomtoets

Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
A1	Is er een mogelijkheid voor een tweede afnamemoment van een inhoudelijk en psychometrische vergelijkbare variant van de in de eerste ronde afgenomen doorstroomtoets (voor leerlingen die de eerste afname zijn verhinderd)?	ja/nee
A2.1	Staat in het toetsreglement/de verantwoordingsdocumenten vermeld hoe de directeur de leerlingen voor de toets moet aanmelden?	ja/nee
A2.2	Staat in het toetsreglement/de verantwoordingsdocumenten vermeld welke hulpmiddelen leerlingen mogen gebruiken?	ja/nee
A2.3	Worden in het toetsreglement/de verantwoordingsdocumenten de voorwaarden aan de (wijze van) geheimhouding van de toetsopgaven door zowel de toetsaanbieder als de deelnemende scholen beschreven?	ja/nee
A2.4	Worden in het toetsreglement/de verantwoordingsdocumenten de voorwaarden aan de (wijze van) waarop toezicht op de leerlingen wordt gehouden beschreven?	ja/nee
A2.5	Worden in het toetsreglement/de verantwoordingsdocumenten de eisen beschreven waar de ICT omgeving (computerconfiguratie) op de school aan moet voldoen om de digitale toets af te kunnen nemen?	ja/nee/n.v.t.

Toelichting A2.1 t/m A2.5:

In de verantwoordingsdocumenten voor de afnemer is de volgende informatie gegeven:

- A2.1: Hoe dient de directeur de leerlingen voor de toets aan te melden?
- A2.2: Welke hulpmiddelen mogen de leerlingen gebruiken?
- A2.3: Welke voorwaarden worden er gesteld aan de (wijze van) geheimhouding van de toetsopgaven door de toetsaanbieder, de schooldirecteur en de afnemers?
- A2.4: Welke voorwaarden zijn er verbonden aan de (wijze) waarop toezicht op de leerlingen moet worden gehouden?
- A2.5: Wat zijn de eisen waar de ICT omgeving (computerconfiguratie) op de school aan moet voldoen om de digitale toets af te kunnen nemen?

Bij het evalueren van deze vragen kijkt de adviseur niet alleen of deze informatie er is, maar ook of deze duidelijk is en geen tegenstrijdigheden bevat. Mocht achteraf blijken dat in de praktijk de gegeven informatie dusdanig verwarrend zijn dat dit een negatieve impact heeft op de afname, dan zal dit in de jaarlijkse evaluatie van de toets een negatief oordeel opleveren.

Beslisregel: Om deze categorie van kwaliteitseisen met een voldoende te kunnen afsluiten, dienen alle vragen met JA te worden beantwoord.

4.4. Beveiligingsaspecten

Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
RB1	Heeft de toetsaanbieder een toetsveiligheidsplan opgesteld met daarin opgenomen een risicoanalyse en een PDCA cyclus voor het beschermen van de beschikbaarheid, integriteit en vertrouwelijkheid van informatie (BIV)?	ja/nee
RB1.1	Zijn de functies en verantwoordelijkheden van de functionarissen van de toetsaanbieder die betrokken zijn bij de inhoud en de beveiliging van de doorstroomtoets beschreven?	ja/nee
RB1.2	Hebben de functionarissen van de toetsaanbieder die betrokken zijn bij de inhoud en de beveiliging van de doorstroomtoets aantoonbaar een geheimhoudingsverklaring ondertekend?	ja/nee
RB1.3	Beschermt de toetsaanbieder het toetsconstructieproces door middel van een PDCA cyclus van beveiligingsprocedures, inclusief aantoonbare maatregelen ter voorkoming van het voortijdig uitlekken van de toetsinhoud?	ja/nee
RB1.4	Is voor het toetsafnameproces een PDCA cyclus van beveiligingsmaatregelen geïmplementeerd, inclusief aantoonbare maatregelen ter voorkoming van het voortijdig uitlekken van de toetsinhoud?	ja/nee

Code	Vragen	Mogelijke antwoorden
RB1.5	Zijn de fysieke en elektronische locaties waar de toetsaanbieder examens, systemen en materialen bewaart, aantoonbaar voldoende beveiligd?	ja/nee
RB1.6	Is de digitale en fysieke informatie omtrent de examenketen zowel tijdens opslag als verzending aantoonbaar voldoende beveiligd?	ja/nee
RB1.7	Verzorgt de toetsaanbieder periodiek training over beveiligingsbewustheid aan haar functionarissen die betrokken zijn bij de inhoud en de beveiliging van de doorstroomtoets?	ja/nee
RB2	Waarborgt de toetsaanbieder, met behulp van passende technische en organisatorische beveiligingsmaatregelen, de privacy en persoonsgegevens van de afnemer, leerling en ouder conform de Algemene Verordening Gegevensbescherming (AVG)?	ja/nee
RB3	Heeft de toetsaanbieder een gedegen en compleet actieplan voor het omgaan met diefstal van (delen van) de toetsinhoud en voor het doen verwijderen van internetpagina's die (delen van) de toetsinhoud onthullen?	ja/nee
RB4	Heeft de toetsaanbieder een gedegen en compleet fraudepreventieplan opgesteld met daarin opgenomen een PDCA cyclus voor de beveiliging van het toetsproces en de toetsinhoud?	ja/nee
RB4.1	Zijn er gedegen en complete procedures voor het ontdekken en evalueren van verdachte toetsresultaten en wijzigingen in kenmerken van toetsonderdelen en toetsscores in de loop van de tijd?	ja/nee
RB4.2	Controleert de toetsaanbieder regelmatig het internet en andere media op het bekend raken van (delen van) de inhoud van de toets of andere niet-openbare informatie over de doorstroomtoets?	ja/nee
RB5	Voorziet de toetsaanbieder in een terugvaloptie, die in geval van kleine incidenten op leerling-, klas- of schoolniveau volgens een bijbehorend calamiteitenplan kan worden ingezet?	ja/nee

Toelichting RB1 t/m RB4.2:

Indien de toetsaanbieder ISO 27001 is gecertificeerd, volstaat het om naar dit certificaat en het achterliggende Information Security Management System (ISMS) inclusief een verklaring van toepasbaarheid (c.q. de scope van het certificaat) te verwijzen.

In alle andere situaties moet de aanbieder overzichtelijk per kwaliteitseis aangeven hoe daar aan is voldaan. Een voorbeeld daarvan is de verwijzing naar een verwerkersovereenkomst, waar op de kwaliteitseisen RB1 t/m RB4.2 wordt ingegaan. Daarbij dient bewijslast te worden overlegd van de bij BI.2, BI.3 en BI.4 gevraagde documenten, processen, contracten en protocollen.

De verantwoordelijkheid voor het goed gebruik van de toets en de daarmee verzamelde informatie in de dagelijkse praktijk ligt bij de school / de eindgebruiker zelf.

Toelichting RB5: De toetsaanbieder is zelf verantwoordelijk om te voorzien in een terugvaloptie in geval van een beperkte calamiteit met alleen de eigen doorstroomtoets. De wijze waarop hierin wordt voorzien kan verschillen per aanbieder en per toetsvorm. De procedure is beschreven in een calamiteitenplan. In geval van een (grootschalige) calamiteit met één of meerdere doorstroomtoetsen, kunnen de getroffen scholen gebruik maken van de door de overheid beschikbaar te stellen calamiteitentoets, zodat de deelnemende scholen ten allen tijde aan hun wettelijke taak voor het afnemen van een doorstroomtoets kunnen blijven voldoen.

Beslisregel: Om deze categorie van kwaliteitseisen met een voldoende te kunnen afsluiten, dienen alle vragen met JA te worden beantwoord.

5. Psychometrische aspecten

5.1. Inleiding

De toetsaanbieder dient nieuw geconstrueerde items te pretesten. Dit om de psychometrische kwaliteit van de items te schatten, om van daaruit te kunnen bepalen welke items mogen worden gebruikt bij de definitieve toetssamenstelling.

De psychometrische kwaliteitsbeoordeling van de doorstroomtoets door een adviseur in opdracht van het CvTE bestaat uit de volgende stappen:

1. Pretestprocedure voor het valideren van nieuw geconstrueerde items, uit te voeren door de toetsaanbieder en bestaande uit:
 - a. Steekproefkader en samenstelling steekproef (paragraaf 5.2.1);
 - b. Kalibratie en kwaliteit van items (paragraaf 5.2.2);
 - c. Toetssamenstelling na pretest (paragraaf 5.2.3).
2. Landelijke item gerelateerde (toetsoverstijgende) normering, uit te voeren door een adviseur in opdracht van het CvTE, zoals beschreven in de separate Regeling Beoordelingsnormen Doorstroomtoetsen po, en bestaande uit:

- a. Het door de toetsaanbieder aanleveren van de ruwe toetsscores, waarna een adviseur deze controleert en analyseert;
 - b. Kalibratie van de gezamenlijk ankeritems;
 - c. Vaststellen van cesuren van referentieniveaus en toetsadviezen;
 - d. Het door de toetsaanbieder vaststellen van de behaalde referentieniveaus en het toetsadvies per leerling.
3. Rapportage achteraf van de kwaliteit en het functioneren van de doorstroomtoets, samen te stellen door de toetsaanbieder en bestaande uit:
 - a. Zes tabellen zoals beschreven in paragraaf 5.3, te verwerken tot één rapportage;
 - b. Weergave van de meest recente data uit Toelatings- en Doorstroomonderzoek. Dit onderzoek wordt centraal uitgevoerd voor het CvTE door een adviseur.

In het geval van een MST, geldt dat voor wat betreft de psychometrische kwaliteitseisen uit hoofdstuk 5 het CvTE per kwaliteitseis mag beoordelen of en welke van de eisen voor een lineaire doorstroomtoets en/of en welke van de eisen voor een computergestuurde adaptieve doorstroomtoets op itemniveau (CAT) van toepassing zijn.

5.2. Pretestprocedure

De pretestprocedure heeft tot doel om de bruikbaarheid van items te bepalen om vervolgens vast te kunnen stellen welke items in aanmerking komen voor opname in de definitief samen te stellen varianten van de doorstroomtoets. Bruikbaarheid betekent dat de items voldoende onderscheidend zijn (d.w.z. een voldoende positieve a-parameter hebben en niet te moeilijk en niet te makkelijk zijn (d.w.z. een representatieve b-parameter hebben). Daarnaast mag de doorstroomtoets ook makkelijke en moeilijke opgaven bevatten om zo goed onderscheid te kunnen maken aan de buitenranden van de vaardigheidsverdeling en rekening houdend met dat de te onderscheiden doorstroomniveaus liggen enkele standaarddeviaties uit elkaar, en is dus ook een goede spreiding in moeilijkheid nodig.

De pretest kan op twee manieren worden uitgevoerd:

1. (eerste afname van items los van de operationele afname):

De toetsaanbieder biedt scholen een vrijwillige proeftoets aan, waarin de nieuw geconstrueerde items zijn opgenomen.

 - Doel A (basis-kwaliteitscontrole):

In de meeste gevallen zal het doel van de proeftoets zijn om de kwaliteit en werking van de items in de praktijk te onderzoeken. Dit betreft zowel het – in een minder high-stakes situatie – onderzoeken van de moeilijkheidsgraad en het onderscheidend vermogen van de items, als het onderzoeken van de kwaliteit van de afleiders. Na de proeftoets selecteert de toetsaanbieder de items die in aanmerking komen voor gebruik in een operationele setting. De gegevens worden alleen gebruikt om items te selecteren. Voor parameterschattingen na de afname voor de vaardigheidsschattingen ten behoeve van het leerlingrapport worden deze data niet gebruikt. Voor dit eerste scenario gelden minder stringente eisen voor de steekproefgrootte en het steekproefkader.
 - Doel B (additionele dataverzameling ten behoeve van itemparameterschattingen):

Het is ook mogelijk een pretest te gebruiken om data te verzamelen die wel gebruikt wordt voor itemparameterschattingen ten behoeve van de vaardigheidsschattingen tijdens de afname. Dat is vooral relevant wanneer de afname niet zo ingericht kan zijn dat van te voren bepaald kan worden dat de opgaven door voldoende leerlingen³ gemaakt zullen worden. Bij een MST of een lineaire toets is dat vaak wel mogelijk, echter bij een CAT kan dat lastiger zijn. Zeker extreme opgaven worden beperkt geselecteerd door het algoritme. Ook als een CAT of MST relatief weinig respondenten heeft kunnen er relatief veel opgaven zijn met minder dan 500–1.000 observaties. Als deze opgaven bij een leerling in de afname zitten, dan moeten er ook voor deze opgaven voldoende gegevens zijn om de parameters te kunnen schatten. Onder voorwaarden die beschreven worden onder het steekproefkader kan de data van de pretest ook gebruikt worden om bij de afname deze opgaven te schalen. Deze voorwaarden zijn strenger dan bij Doel A. Het is ook mogelijk om doel B na te streven bij MST of lineaire toetsen, waarbij dan ook de strengere voorwaarden gelden.
2. (eerste afname binnen de operationele afname; zaaien, d.w.z.):

Opgaven uitproberen als onderdeel van een werkelijke afname waardoor de opgave in dezelfde omstandigheden gemaakt wordt als waarin ze toegepast moeten worden): De toetsaanbieder zaait de nieuwe items in de operationele versie van de doorstroomtoets. Wanneer uit de pretestprocedure blijkt dat de nieuwe items bruikbaar zijn, mag de toetsaanbieder deze direct mee laten tellen voor de berekening van het toetsadvies en/of de score op de referentieniveaus. Voor een leerling mag het niet herkenbaar zijn of een vraag wel of niet meetelt. Voor dit tweede scenario gelden

³ Voor de aantallen wat als voldoende gezien wordt, zie *Toelichting N1.1a*

strengere eisen voor het minimaal vereiste aantal observaties. Als aan alle eisen voldaan wordt, maar alleen te weinig observaties per opgave waargenomen worden, hoeft dit niet tot een onvoldoende beoordeling te leiden: het is dan mogelijk de resultaten verkregen onder scenario 2 te gebruiken voor doel A van scenario 1 mits het aantal observaties minimaal gelijk is als in dat geval vereist is.

Tevens geldt dat voor de berekening van het toetsadvies en de berekening van de score op de referentieniveaus alleen gebruik gemaakt mag worden van gekalibreerde (pretest)items die onderwijskundig en psychometrisch goed functioneren. Wanneer de toetsaanbieder er toch voor kiest om, om andere redenen, minder goed functionerende items mee te laten tellen, dan dient de toetsaanbieder dit te beargumenteren in de wetenschappelijke verantwoording van de doorstroomtoets.

De pretestprocedure bestaat in beide scenario's uit drie stappen: 1) Steekproefkader en samenstelling steekproef vaststellen (§ 5.2.1), 2) Kalibratie en kwaliteit van items vaststellen met de verzamelde pretestdata (§ 5.2.2), en 3) Definitieve doorstroomtoets samenstellen (§ 5.2.3).

5.2.1. Steekproefkader en samenstelling steekproef

De onderwijskundig positief beoordeelde nieuw geconstrueerde items worden gepretest volgens de kwaliteitseisen voor het steekproefkader en de samenstelling van de steekproef.

Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
N1.1	Is de gebruikte steekproef van voldoende kwaliteit, gezien het doel van de pretest?	ja/nee
N1.1a	Is de steekproef groot genoeg?	ja/nee
N1.1b	Is de steekproef relevant?	ja/nee
N1.2	Heeft de pretest plaatsgevonden conform het intern dataverzamelingsdesign of het gecombineerd dataverzamelingsdesign; en is het dataverzamelingsdesign van de pretest adequaat?	ja/nee
N1.3	Is, in het geval van open opgaven in een pretest, de beoordelaarsovereenstemming gegeven?	ja/nee/n.v.t.

Toelichting N1.1:

De beantwoording van deze vraag is afhankelijk van het doel van de steekproef. Hierboven is al verwezen naar de twee scenario's die van toepassing zijn. De vraag naar de kwaliteit van de steekproef is verder opgedeeld naar de grote en de relevantie van de steekproef. Er is automatisch voldaan aan kwaliteitseis N1.1, wanneer is voldaan aan de kwaliteitseisen N1.1a en N1.1b.

Toelichting N1.1a:

Als onder *scenario 1 doel A* wordt nagestreefd (bepalen of een opgave van voldoende kwaliteit is om opgenomen te mogen worden in de operationele doorstroomtoets) is het benodigde aantal leerlingen beperkt. Uit een relatief beperkt aantal observaties valt immers te achterhalen of een opgave goed genoeg functioneert om opgenomen te worden in de afname. De focus moet hierbij vooral liggen op het onderscheidend vermogen van de opgave en de werking van de afleiders in het geval van een meerkeuze-opgave. Om de kans te verkleinen dat het onderscheidend vermogen bij de afname negatief blijkt te zijn is het verstandig een redelijke ondergrens van de Rit-waarde of de geschatte a -parameter te kiezen. De aanbieder is hier vrij in, maar dient wel aan te geven welk criterium gehanteerd is (en waarom). In de situatie dat een bestaande itembank wordt uitgebreid c.q. wordt ververs, wordt het aantal observaties als voldoende gezien wanneer iedere opgave door minstens 200 leerlingen gemaakt is. In de situatie dat er een geheel nieuwe doorstroomtoets wordt gemaakt, geldt een minimum van 400 leerlingen. Uiteraard geven meer gegevens meer zekerheid. Wanneer de steekproef op klassenniveau is samengesteld, dient de aanbieder op adequate wijze rekening te houden met de intraklasse-correlatiecoëfficiënt (ICC of ρ)⁴ en moet de aanbieder de steekproefgrootte ophogen op basis van de hoogte van de ρ en het aantal leerlingen in de klas dat dezelfde opgaven maakt (m). De ρ kan worden berekend aan de hand van de variantie binnen de klassen (sw^2) en de variantie tussen de klassen (sb^2) met behulp van de formule: $\rho = sb^2 / (sb^2 + sw^2)$. Vervolgens kan de vereiste steekproefgrootte op klassenniveau worden bepaald met de formule: $N_{trekking} = N_{doel} (1 + \rho (m - 1))$.⁵

⁴ H.Barratt, M.Kirwan, S. Shantikumar, *Clustered data – effects on sample size and approaches to analysis*, HealthKnowledge 2018 (in revisie).

⁵ Uitgewerkt in een rekenvoorbeeld: Wanneer de steekproef 200 leerlingen moet bevatten, is dat de waarde van N_{doel} . In het geval vervolgens een schoolklas van 24 leerlingen wordt opgedeeld in twee groepen (c.q. klassen), dan is m in de berekening van de steekproef voor het aantal observaties per item 12. Bij een fictieve ρ van 0,07 is de benodigde steekproefgrootte op klassenniveau dan: $N_{trekking} = 200 * (1 + (0,07 * (12 - 1))) = 354$ leerlingen.



Wanneer onder *scenario 1 doel B* wordt nagestreefd, dan dient het aantal observaties hoger te zijn. Op het moment van itemkalibratie t.b.v. de parameterschattingen voor de rapportage dienen dezelfde aantallen behaald te worden als gespecificeerd staan onder scenario 2.

Onder *scenario 2* worden de opgaven voor het eerst afgenomen direct onder operationele afnameconditie. We gaan er hierbij van uit dat deze opgaven van tevoren onderwijskundig als kwalitatief voldoende gezien worden. Als de psychometrische kenmerken dan ook goed zijn, dan onderscheiden deze opgaven zich noch in onderwijskundig opzicht, noch in psychometrisch opzicht van de andere opgaven in de toets. Omdat leerlingen er tijd en moeite in steken is het terecht als deze items dan meetellen voor de rapportage. In dat geval kunnen deze items tijdens de campagne bijgeschaald worden. Om items mee te laten tellen voor de terugrapportage naar leerling moeten de itemparameters goed geschat worden. Daartoe moeten er voldoende observaties per item zijn. Voor dichotoom gescoorde items geldt bij een 2PL itemresponsemodel een minimum aantal van 1.000 observaties. Voor polytoom gescoorde items geldt bij een 2PL itemresponsemodel een minimum aantal van 1.200 observaties. Aanbieders mogen van deze aantallen afwijken, mits dit gedegen en volledig wordt onderbouwd.

Het is mogelijk dat de aanbieder ook onder scenario 2 alleen doel A van scenario 1 nastreeft. Het minimaal vereiste aantal observaties is dan gelijk aan dat bij doel A van scenario 1. Dan tellen de opgaven niet mee in het leerlingrapport.

Voor alle items die meetellen voor de rapportage dienen de bovengenoemde aantallen behaald te worden. Bij opgaven die bijvoorbeeld in een CAT of MST te weinig door het algoritme geselecteerd worden, worden de aantallen niet altijd tijdens de afname gehaald. In die gevallen mogen de gegevens van de pretest ook gebruikt worden voor de kalibratie om die itemparameters te schatten. De genoemde aantallen observaties per opgave kunnen dan gehaald worden met de som van de pretest-observaties en de afname-observaties. Let wel dat dan ook aan andere steekproefvragen voldaan moet zijn aangaande representativiteit.

Gekloonde opgaven zijn in beginsel nieuwe opgaven, en worden zo ook beoordeeld. Wanneer er sprake is van (zeer) kleine wijzigingen, zoals de aanpassing van een munteenheid (van euro naar dollar), of de aanpassing van een naam (bijvoorbeeld ten behoeve van inclusie), waarbij de opgave inhoudelijk hetzelfde blijft, dan kunnen dergelijke items als onveranderd gezien worden. De toetsaanbieder dient dan wel duidelijk aan te geven welke (inhoudelijke) aanpassing gedaan is, bij voorkeur met een rationale waaruit blijkt waarom de aanpassing geen impact heeft.

Wanneer de gekloonde opgave in termen van relevantie en/of objectiviteit verandert ten opzichte van het originele item, dan is er wel sprake van een nieuwe opgave. Meer specifiek is er sprake van een nieuwe opgave wanneer:

- de gekloonde opgave een relatie heeft met een ander of gewijzigd leerdoel;
- de gekloonde opgave naar verwachting een andere of gewijzigde moeilijkheidsgraad heeft;
- de gekloonde opgave naar een ander of gewijzigd modelantwoord leidt.

Wanneer eerder al data verzameld zijn bij het item in een andere vorm dient ook bij kleine wijzigingen het gekloonde item een nieuw label te krijgen.

Toelichting N1.1b:

De steekproef dient relevant te zijn voor de toepassing. Dit betekent dat de steekproef van de pretest overeenkomstig dient te zijn met de populatie die de toets van de aanbieder regulier afneemt.

Voor scenario 2, het zaaien in de afname, is dat relatief makkelijk in te richten door de zaai-items verstandig op scholen in te zetten. Deze steekproefprocedure dient door de aanbieder beschreven te worden. Een ander voordeel bij scenario 2 is dat de afnamecondities per definitie identiek zijn aan reguliere afnamecondities zolang de items niet specifiek als zaai-items te herkennen zijn.

In het geval van scenario 1 dient een aparte steekproef getrokken te worden. De eisen aan deze steekproef en de afnamecondities verschillen voor doel A en doel B bij dit scenario. Onder doel A is het aan te raden de steekproef sterk te doen lijken op toets-populatie. Een variatie aan scholen is hierbij aan te raden, waarbij rekening gehouden dient te worden met de variabelen in onderstaande tabel. Het is aan te bevelen minstens twee niveaus van iedere variabele mee te nemen. Voor de variabele schoolweging geldt een ondergrens van drie niveaus. Als scenario 1 doel B nagestreefd wordt, dan wordt verwacht dat de afname zoveel mogelijk onder dezelfde (waarschijnlijk high-stakes) condities plaats vindt. Er dient door de aanbieder beschreven te worden hoe dat zoveel mogelijk bereikt wordt. Tevens dient aangetoond te worden dat de verdeling van de steekproef goed gelijkend is aan de toets-populatie. Daarvoor mag de toets-populatie van het voorgaande afnamejaar als referentie en steekproefkader gebruikt worden. Deze populatie en steekproef dienen dan beschreven



te worden op basis van de variabelen en niveaus zoals beschreven in de onderstaande tabel.

Variabele	#N	Welke niveaus (NB: #N = aantal niveaus)
regio	4	noord, zuid, oost, west
urbanisatiegraad	3	G4 (de vier grootste gemeenten), G5-G37 (de 5 tot 37 grootste steden), >G37 (de overige steden)
schoolgrootte	3	aantal leerlingen: <100, 100–300, >300 leerlingen
schoolweging	5	gewichtscategorieën: <23 / 23–26,99 / 27–32,99 / 33–36,99 / ≥ 37

Onder *scenario 1* wordt een positief antwoord op deze vraag gegeven wanneer aan de volgende vijf voorwaarden is voldaan:

- 1) Op alle in de tabel genoemde variabelen zitten er op minimaal twee niveaus per variabele en op minimaal drie niveaus van de variabele schoolweging gegevens in de steekproef;
- 2) Er is aangegeven hoe de steekproef getrokken is;
- 3) Er is aangegeven onder welke omstandigheden de data verzameld zijn;
- 4) Er is aangegeven hoe de steekproef zich verhoudt tot de landelijke populatiewaarden van leerlingen in groep 8 met betrekking tot de in de tabel genoemde niveaus per variabele;
- 5) De steekproefgegevens over S(B)O scholen worden aanvullend separaat gerapporteerd en verantwoord in lijn met de eigen populatie uit voorafgaande schooljaren.

Als bij drie (3) deze omstandigheden afwijken van de werkelijke afname wordt aangegeven welke impact verwacht wordt en hoe daar in de procedure van de selectie van opgaven voor de operationele afname rekening mee is gehouden.

Onder *scenario 2* wordt een positief antwoord wordt op deze vraag gegeven wanneer aan de volgende twee voorwaarden is voldaan:

- 1) De aanbieder toont aan dat de populatie voor de zaai-items niet structureel anders is dan de populatie(s) van de rest van de opgaven.
- 2) Bij de observaties is een steekproefkader (c.q. een design van hoe de toetsboekjes, de CAT toetsversies en/of de MST toetspaden met zaai-items en ankeropgaven worden ingezet) gegeven met een beschrijving van de procedure waarmee de steekproef van observaties tot stand is gekomen;

Toelichting N1.2:

In Scenario 1 doel A, wanneer de pretest alleen een rol speelt in het al dan niet selecteren van de opgaven, en de bij de normering (definitieve) gebruikte parameters bepaald worden op basis van de operationele afname, is deze eis niet van toepassing. De kwaliteit van de pretest wordt dan alleen bepaald op basis van N1.1.

Wanneer er sprake is van *Scenario 2 of van Scenario 1, doel B* is het van belang dat de parameters per vaardigheid op dezelfde schaal liggen. Dat geldt dus ook voor de zaai-items. In het geval van een CAT dienen de parameters van de opgaven die de itembank vormen op dezelfde schaal te liggen. Om dat voor elkaar te krijgen is het van belang dat er sprake is van een afnamedesign waarbij dit mogelijk is. Wanneer de pretest gedaan wordt door middel van 'zaai-opgaven' in de afname dan wordt aangegeven hoe dat 'zaai-design' er uitziet. Dit design is meegegeven in de technische verantwoording van de pretest en bevat ieder geval informatie over:

- het totaal aantal boekjes ('booklets');
- aantal observaties per boekje;
- het aantal items dat tussen de opgavenboekjes onderling overlapt;
- het aantal items dat de pretest verbindt met de bestaande schaal.

Met behulp van deze informatie kan worden bepaald of de boekjes aantoonbaar voldoende 'gelinked' zijn. Dat betekent dat er voldoende overlap in observaties tussen de verschillende items en boekjes moet zijn. Daarvan is sprake wanneer de aanbieder laat zien dat het plaatsen van de itemparameters van alle opgaven op dezelfde schaal als waar de meting over gaat mogelijk is, meestal door middel van een verbonden design.

Wanneer de pretestdata zijn verzameld met een CAT is het niet zinvol om het hele onvolledige design weer te geven, omdat iedere leerling dan in principe een unieke verzameling items maakt. Het dient wel duidelijk te zijn op welke gegevens de linking en normering gebaseerd zijn. Bij een MST of een lineaire afname is het meeleveren van het design wel zinvol.

In het geval van een CAT is, net als bij de andere toetsvormen, aangegeven hoeveel zaai-items een kandidaat kan krijgen, en hoe die opgaven toegewezen worden.

Ook moet duidelijk zijn op welke posities de verschillende opgaven ingezet zijn, zodat door de toetsaanbieder op gedegen en complete wijze beargumenteerd kan worden op welke wijze er rekening

is gehouden met vermoeidheids- en/of volgorde effecten.

Toelichting N1.3: De kwaliteit van de niet geautomatiseerd gescoorde open opgaven wordt ook beïnvloed door de beoordelaarsovereenstemming. In de pretest moet er bij niet geautomatiseerd gescoorde open opgaven daarom ook beoordelaarsovereenstemmingsonderzoek worden uitgevoerd. In het geval van Scenario 1 doel A moet aangegeven worden op basis van welke criterium items vanwege een lage beoordelaarsovereenstemming niet meer opgenomen worden, en wanneer opgaven wel opgenomen kunnen worden. Bij scenario 2 en bij scenario 1 doel B moet aangegeven worden wat de impact is op de geschatte parameters.

Beslisregel: Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, dienen alle vragen met JA te worden beantwoord.

5.2.2. Kalibratie en kwaliteit van items

Na het verzamelen van de observaties voert de toetsaanbieder een kalibratie uit met de verzamelde testdata. In het geval van eerder afgenomen opgaven (bij een itembank of bij pretestgegevens) kan er een schaal gekalibreerd worden die op basis van de afnamegegevens bestendig moet worden. Op basis van de afnamegegevens kunnen de (goed functionerende) zaai-items op de schaal geplaatst worden. Bij de beoordeling wordt gekeken naar wat er in het verleden (vorige afname) gedaan is en wordt geëvalueerd wat de plannen zijn voor bij de afname. Wanneer een aanbieder voor het eerst een afname heeft, is de toelating voorlopig en kan deze na de eerste afname bevestigd worden op basis van de resultaten bij deze afname, vergelijkbaar met hoe aanbieders met eerdere afnames vooraf beoordeeld worden. In de kalibratie schat de toetsaanbieder de itemparameters, zodat de kwaliteit van de met behulp van scenario 1 of scenario 2 gepreteste items kan worden vastgesteld.

Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
KA1	Zijn de itemparameters van de nieuwe items geschat met een bij de data passend psychometrisch model?	ja/nee
KA2	Is er op gedegen en complete wijze beschreven dat de itemparameters van de geselecteerde gekalibreerde items toepasbaar zijn?	ja/nee
KA3	Worden beslisregels geëxpliciteerd om dekking van het inhoudelijke domein te waarborgen per individuele afname?	ja/nee/n.v.t.
KA4	Is er een schatting gemaakt van de populatieparameters?	ja/nee
KA5	Zijn de metingen van de individuele leerlingen voldoende betrouwbaar om uitspraken op individueel niveau te kunnen doen?	ja/nee
KA6	Is er aannemelijk gemaakt dat het gebruikte model past bij de data?	ja/nee
KA7	Bij gebruik van een 2PL IRT model, is er een indicatie gegeven van de lokale betrouwbaarheid van de nieuwe items?	ja/nee

Toelichting KA1:

De toetsaanbieder schat van alle, in scenario 1, doel B of in scenario 2 gepreteste items de moeilijkheidsgraad (p-waarde of een vergelijkbare parameter), het discriminerend vermogen (Rit-waarde en Rar-waarde of vergelijkbare parameters) en, in het geval van meerkeuzevragen de kwaliteit van de afleiders (a-waarde of een vergelijkbare parameter) in een 2PL itemresponsemodel en met gebruik van hiertoe geëquipeerde software.

De toetsaanbieder selecteert vervolgens de items die psychometrisch gezien van voldoende kwaliteit zijn om te kunnen gebruiken in de samen te stellen doorstroomtoets. Dat betekent dat de toetsaanbieder op gedegen en complete wijze onderbouwt dat van de geselecteerde items:

- de moeilijkheidsgraad passend is voor het meetdoel en de doelgroep (d.w.z. voor lineaire toetsen vergelijkbaar met een p-waarde tussen de gokkans en 0.95 en voor adaptieve toetsen passend bij de moeilijkheidsgraad van het betreffende design en/of toetspad);
- het discriminerend vermogen vergelijkbaar is met een Rit-waarde > 0.15 , rekening houdend met de beperkte vaardigheidsrange van de leerlingen in het geval van een adaptieve toets;
- de kwaliteit van de afleiders, in het geval van meerkeuzevragen, vergelijkbaar is met een a-waarde > 0.05 ;
- de interbeoordelaarsovereenstemming, in het geval van open vragen, vergelijkbaar is met een interbeoordelaarsovereenstemmingscoëfficiënt > 0.70 .

De toetsaanbieder verantwoordt op gedegen en complete wijze de keuze voor de geselecteerde items die in aanmerking komen voor gebruik binnen de samen te stellen doorstroomtoets.



In het geval van scenario 1, doel A dient de aanbieder te verantwoorden op basis van welke criteria opgaven toegelaten worden voor de operationele afname.

Toelichting KA2:

De toetsaanbieder beschrijft op gedegen en complete wijze dat de itemparameters van de geselecteerde items toepasbaar zijn. Dit wordt gedaan door (1) aan te tonen dat de itemparameters tussen verschillende afnamen of boekjes niet veranderen, en (2) een bij de data en een bij de populatie passende DIF analyse (differential item functioning) op itemniveau uit te voeren op bijvoorbeeld de achtergrondvariabele regio, waarbij de itemparameterschattingen uit de verschillende boekjes in de pretest worden vergeleken. Wanneer er sprake is van items met DIF, dient de toetsaanbieder aan te tonen dat dit in overeenstemming is met de verwachting op basis van de relevante literatuur.⁶ Het ontbreken van de genoemde DIF-analyse kan resulteren in een onvoldoende voor deze kwaliteitseis.

Toelichting KA3:

De theorie achter adaptief toetsen gaat er van uit dat een efficiëntere schatting van het vaardigheidsniveau van de leerling kan worden verkregen als het aanbod van de items steeds wordt aangepast aan de antwoorden van de leerling op voorafgaand items. Derhalve dienen van dit type doorstroomtoets de beslisregels of de algoritmes voor de samenstelling van de toets te zijn geëxpliciteerd. De toetsaanbieder dient op gedegen en complete wijze te verantwoorden: (1) hoe de toets wordt gestart, (2) hoe de keuze voor een volgend item wordt gemaakt, en (3) wanneer de toets wordt beëindigd. Naast dat de beslisregels geëxpliciteerd dienen te zijn, is het belangrijk dat tevens beslisregels geëxpliciteerd zijn om dekking van het inhoudelijke domein te waarborgen per individuele afname. Dit houdt in dat indien er bijvoorbeeld vier verschillende domeinen van rekenen worden bevraagd, de beslisregels van de adaptieve toets ervoor dienen te zorgen dat ook uit alle domeinen opgaven worden afgenomen. Wanneer of de startprocedure, of de selectieprocedure of de stopprocedure niet op gedegen en complete wijze is beschreven, kan dit resulteren in een onvoldoende voor deze kwaliteitseis.

Toelichting KA4:

Bij KA4 wordt een positief oordeel gegeven wanneer de soort verdeling (normale verdeling, of anders) en het gemiddelde, de standaardafwijking en hun schattingsfout (standard error) op de vaardigheidsverdeling worden gerapporteerd, en er wordt verantwoord hoe deze schattingen zijn verkregen. De toetsaanbieder geeft daarbij aan voor welke populatie(s) deze parameters gegeven worden. Dit dient ieder geval gegeven te worden voor de reguliere po-leerlingpopulatie.

Voor doorstroomtoetsen die eerder zijn afgenomen is deze schatting gebaseerd op de daadwerkelijke populatievaardigheid van het voorgaande jaar. Bij een eerste afname is dat niet mogelijk, en zal aangegeven moeten worden welke verwachtingen men hier heeft. Dit gegeven wordt ook verwacht in de jaarlijkse rapportage na de afname, zoals beschreven bij 5.3.

Bij KA4 wordt ook verwacht dat de aanbieder expliciet maakt welke aannames hij doet/gaat doen.

Toelichting KA5:

De meting van de vaardigheid van een leerling moet voldoende betrouwbaar worden uitgevoerd. Dit betekent dat het van belang is dat de schattingsfout (standard error) van iedere geschatte vaardigheid beperkt is. Deze schattingsfout zal in de praktijk (waarschijnlijk) niet voor iedere leerling even groot zijn. Daarom dient aangegeven te worden wat de verdeling van de schattingsfouten is. Om de grootte van de schattingsfout op waarde te kunnen schatten zal de standaardafwijking zoals deze gegeven wordt in KA4 als referentie dienen. Gezien het belang van de classificaties, moet aangetoond worden dat de schattingsfout rond ieder van de cesuurpunten van de referentieniveauschalen correspondeert met een klassieke betrouwbaarheidscoëfficiënt $> 0,70$.

De eis van de minimale betrouwbaarheid (de manier om de schattingsfout van de persoonsparameters in te schatten) wordt op drie manieren beoordeeld:

- Is er een gedegen en complete beschrijving van de procedure, waaruit blijkt dat de betrouwbaarheid op de juiste manier is berekend?
- Is de hoogte van de betrouwbaarheid rond de cesuurpunten van de referentieniveaus groter dan 0,70?
- Is de hoogte van de betrouwbaarheid van de schaal waarop de toetsadviezen gegeven worden minimaal 0,80?

Toelichting KA6:

Alle gebruikte modellen werken met aannames en voorwaarden waaraan voldaan moet worden, wil

⁶ Zie bijvoorbeeld: Bechger, Exploring Differential Items Functioning with Dexter, 2022.

het model bruikbaar zijn. In de landelijke normering zal het 2PL gebruikt worden zodat voldoende moet worden aangetoond dat de verkregen data voldoende past bij de voorwaarden die bruikbaarheid van het model vereist. Dit betreft de veronderstelling van (1) unidimensionaliteit, (2) lokaal stochastische onafhankelijkheid, en (3) de passing van de data bij de gebruikte parameterschattingen. Onder unidimensionaliteit wordt een eigenschap van een schaal verstaan wanneer de items in een toets dezelfde vaardigheid meten. Met elke willekeurige subset van items uit de gekalibreerde itembank kan dezelfde onderliggende vaardigheid worden vastgesteld. Ook onderzoek naar DIF wordt hierbij als onderbouwing gezien. Hier zijn diverse maten en onderzoeken mogelijk, waarbij de toetsaanbieder hier met de juiste motivering aangeeft hoe aan de voorwaarden voldaan is. Naast statistische passingsmaten is het ook mogelijk gebruik te maken van bij de data en het psychometrisch model passende robuustheidsmaten.

Toelichting KA7:

Het is van belang dat de aangeboden opgaven voldoende onderscheidend zijn. In het geval van een itembank zullen items met weinig onderscheidend vermogen zelden of nooit aangeboden worden. Deze items vormen daarmee een overschatting van de vulling van de bank in aantallen opgaven. In het geval van een lineaire toets of een MST, is het onwenselijk dat een leerling tijd besteed aan opgaven met een extreem laag gewicht.

Om een positief oordeel op deze vraag te krijgen wordt aangetoond dat alle opgaven een voldoende onderscheidend vermogen hebben. Dat wordt aangetoond door te laten zien dat het onderscheidend vermogen van de opgaven in de bank of toets niet kleiner is dan 10% van het gemiddelde onderscheidend vermogen.

Beslisregel: Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, dienen alle vragen met JA te worden beantwoord.

5.2.3. Toetssamenstelling

Na afronding van de kalibratie van de nieuwe items, stelt de toetsaanbieder de definitieve doorstroomtoets samen.

Kwaliteitseisen

Code	Vragen	Mogelijke antwoorden
T1	Bevat de doorstroomtoets voor de terreinen Nederlandse taal en rekenen in ieder geval vragen met een moeilijkheidsgraad tussen de cesuur van referentieniveau 1F en 2F/1S?	ja/nee
T2	Bevat de doorstroomtoets het door het CvTE voorgeschreven minimaal aantal gezamenlijke ankeritems voor de wettelijk verplichte terreinen Nederlandse taal en rekenen? Is de volledige subset van gezamenlijke ankeritems opgenomen in (de varianten van) de doorstroomtoets? Zijn de gezamenlijke ankeritems op de in de Regeling Beoordelingsnormen Doorstroomtoetsen po beschreven wijze opgenomen in de doorstroomtoets?	ja/nee ja/nee
T3	Bestaat een papieren toets of een lineaire digitale toets jaarlijks vrijwel uit nieuwe items voor de wettelijk verplichte terreinen Nederlandse taal (domeinen Lezen en Taalverzorging) en rekenen, met uitzondering van de items van het gezamenlijk anker?	ja/nee/n.v.t.
T4	Heeft de itembank van een CAT een omvang van minimaal 6x het aantal items van een maximale operationele toetsversie, waarbij jaarlijks tussen de 20% en 30% van de items wordt ververs?	ja/nee/n.v.t.
T4.1	Is er, in het geval van een MST die bestaat uit meerdere lagen van modules (c.q. routes), sprake van een adequate verversingsstrategie a. ... voor niet geselecteerde items? b. ... voor items met discrepanties in parameters in het kalibratiedesign en de voorlaatste operationele fase? c. ... voor items die uitgelekt lijken? d. ... voor items die over jaargangen heen anders functioneren?	ja/nee/n.v.t. ja/nee/n.v.t. ja/nee/n.v.t. ja/nee/n.v.t.

Toelichting T1: Voor zowel een papieren als een digitale lineaire doorstroomtoets geldt dat de toets voor de terreinen Nederlandse taal en rekenen in ieder geval vragen moet bevatten die inhoudelijk aansluiten bij de referentieniveaus 1F en 2F/1S en die een moeilijkheidsgraad tussen de cesuur van referentieniveau 1F en 2F/1S hebben. In geval van CAT is de itemselectie geoptimaliseerd rond de opeenvolgende vaardigheidsschattingen. Er dient te worden voldaan aan de inhoudelijke randvoorwaarden, terwijl de lokale betrouwbaarheid rond de cesuren van de referentieniveaus en de toetsadviescategorieën hoog genoeg is.

Voor een CAT geldt dat de itembank voldoende groot moet zijn. Dat wil zeggen dat er een gesloten itembank is van minimaal zes (6) keer de maximale toetslengte per te toetsen domein.



Toelichting T2a: De door het CvTE toegewezen subset van gezamenlijke ankeritems is opgenomen in (de varianten van) de doorstroomtoets.

Toelichting T2b: De items van de verplicht op te nemen subsets van het gezamenlijk anker moeten op een zodanige wijze over de toets worden verspreid, dat de positie in de toets zo min mogelijk invloed op de prestatie van de leerling heeft en de ankeritems qua lay-out en tekstuele vormgeving niet als opvallend afwijkend in de toets herkenbaar zijn, een en ander zoals gespecificeerd in de Regeling Beoordelingsnormen Doorstroomtoetsen po.

Toelichting T3: Een papieren toets of een lineaire digitale toets bestaat jaarlijks vrijwel uit nieuwe gekalibreerde items voor de wettelijk verplichte domeinen Nederlandse taal en rekenen, met uitzondering van de items van het gezamenlijk anker.

Toelichting T4 / T4.1:

In het geval van een MST op basis van een itembank die bestaat uit meerdere lagen van modules (c.q. routes) of in het geval van een CAT, moet er sprake zijn van een adequate verversingsstrategie, en de verversing moet op itemniveau gedocumenteerd zijn. Daarbij geldt:

- De aanbieder hanteert jaarlijks een adequate verversingsstrategie voor de opgaven van de itembank van een MST of een CAT. De toetsaanbieder geeft op gedegen en complete wijze aan hoeveel procent van de opgaven is ververst.
- Items die uitgelekt lijken, moeten verwijderd worden. De toetsaanbieder geeft op gedegen en complete wijze aan hoeveel items er vermoedelijk zijn uitgelekt en hoeveel van deze items uit de itembank zijn verwijderd.
- Items die over de jaren heen andere itemparameters hebben, moeten verwijderd worden. De toetsaanbieder geeft op gedegen en complete wijze aan hoeveel items er veranderende itemparameters hebben en hoeveel van deze items uit de itembank zijn verwijderd.

De punten b en c kunnen onderzocht worden met een statistische toets voor DIF en/of met een statistische toets voor Item Parameter Drift (IPD).

		Aantal gevonden	Aantal verwijderd uit itembank
b.	Items die uitgelekt lijken		>90%
c.	Items die over jaargangen heen anders functioneren		>90%

De aanbieder verantwoordt op gedegen en complete wijze hoe de aantallen in de kolom 'Aantal gevonden' zijn bepaald. Van de gevonden afwijkingen dient in hetzelfde jaar minimaal 90% aantoonbaar te zijn verwijderd uit de itembank.

Om het gebruik van een adaptieve doorstroomtoets aantrekkelijk te houden, dient kwaliteitseis T4.1 te worden gelezen in termen van: 'het onderbouwen en motiveren van de gemaakte keuzes'.

Beslisregel: Om voor deze categorie van kwaliteitseisen een voldoende te kunnen krijgen, dienen alle van toepassing zijnde vragen met JA te worden beantwoord.

5.3. Rapportage achteraf van de kwaliteit en het functioneren van de doorstroomtoets

De toetsaanbieder verantwoordt jaarlijks na de toetsafname-periode de kwaliteit en het functioneren van de afgenomen doorstroomtoets. De aanbieder stelt daartoe minimaal de volgende zes tabellen samen, conform de inhoud en indeling zoals voorgeschreven door het CvTE. Aansluitend controleert een adviseur in opdracht van het CvTE of de inhoud van de rapportage voldoet aan de vooraf gestelde eisen. De resultaten en de kwaliteit van dit rapport van het voorafgaande jaar worden gebruikt bij de evaluatie in dit kader, dan wel de jaarlijkse her-evaluatie. Als er een nieuwe aanbieder is die dit nog niet op kan leveren naar aanleiding van een vorige afname, speelt dit bij de eerste evaluatie geen rol, maar heeft het wel een groter gewicht bij de jaarlijkse her-evaluatie die volgt na de eerste afname.

De jaarlijkse verantwoording van de kwaliteit en het functioneren van de doorstroomtoets dient uiterlijk op 31 mei beschikbaar te zijn gesteld aan het CvTE en bestaat uit de volgende informatie:

- Een tabel met aantallen opgaven per vaardigheid: aantal wel en aantal niet meegenomen voor de scorerapportage naar de leerlingen wegens (1) een inhoudelijke fout, (2) psychometrisch disfunctioneren, (3) DIF, (4) te weinig waarnemingen. Wanneer er zaai-items ingezet worden kunnen deze in een aparte tabel weergegeven worden. Het leveren van een lijst welk item dit betreft wordt aangeraden.
- Een overzicht (tabel) van de percentages referentieniveaus van de wettelijk verplichte en optionele domeinen van het terrein Nederlandse taal (domein lezen en taalverzorging) en het terrein rekenen



plus een overzicht van de percentages toetsadviezen (de zes categorieën) van dit jaar ten opzichte van het vorige jaar.

3. Een overzicht (tabel) van de relatie tussen de referentieniveaus en de achtergrondvariabelen van de doelgroep.
4. Een overzicht (tabel) van de relatie tussen de toetsadviezen en de achtergrondvariabelen van de doelgroep.
5. Voor wat betreft de eerder afgenomen doorstroomtoets, een overzicht (tabel) van de verdeling van de populatievaardigheid (gemiddelde en standaardafwijking) van in ieder geval de reguliere po-populatie inclusief de schattingsfout (standard error) op de vaardigheidsverdeling, een en ander op dezelfde schaal als de itemparameters.
6. Bij eerder afgenomen opgaven (in eerder jaar of bij pretest): DIF-analyse-gegevens met betrekking tot parameters van de meest recente operationele afname en de eerdere afname(n), een en ander ter onderbouwing van de bij de kwaliteitseisen T4 en T4.1 beoordeelde verversingsstrategie van de aanbieder.

Een overzicht (tabel) van de overeenkomsten tussen de toetsadviezen en de eerder gegeven (voorlopige) schooladviezen, bestaande uit een overzicht van het percentage consistente en niet consistente classificaties, inclusief de globale classificatiefout (c.q. het percentage misclassificaties als functie van de mate van overeenstemming tussen het voorlopig schooladvies en het toetsadvies⁷)

6. Vierjaarlijkse erkenning en jaarlijkse vaststelling door het CvTE

6.1. De vierjaarlijkse erkenning

Een adviseur beoordeelt vierjaarlijks in opdracht van het CvTE de onderwijskundige, organisatorische en psychometrische aspecten van de doorstroomtoets. Na de beoordeling stelt een adviseur een advies op. Dit advies gaat naar het CvTE dat, op basis van het ontvangen advies, een definitieve beslissing neemt over het al dan niet erkennen van de betreffende doorstroomtoets. Wanneer de toets is erkend kan deze toets vier achtereenvolgende jaren worden ingezet. Onderdeel van deze uitgebreide beoordeling is een psychometrische controle van de pretestprocedure én van de operationele testprocedure en -resultaten die jaarlijks worden verzameld in de afnameperiode.

Na de eerste toelating van vier jaar, dienen de toetsaanbieders opnieuw een toelatingsaanvraag in bij het CvTE. Voor deze nieuwe toelatingsaanvraag vindt een herbeoordeling plaats. Voor deze herbeoordeling van de onderwijskundige inhoud, organisatorische aspecten en psychometrische aspecten doet het CvTE eveneens een beroep op onafhankelijke onderwijskundige en vakinhoudelijke experts van een adviseur.

6.2. De jaarlijkse vaststelling

Een adviseur beoordeelt jaarlijks aan de hand van hetzelfde beoordelingskader de kwaliteit van de doorstroomtoets. De aanbieder levert hiervoor bij het CvTE alle gevraagde documenten aan die eveneens voor de vierjaarlijkse erkenning ingediend dienen te worden. De aanbieder verstrekt hierbij een leeswijzer. Hierin staat per document aangegeven op welke punten deze afwijkend is van de eerder verstuurde versie ten tijde van de vierjaarlijkse check of jaarlijkse vaststelling. Vervolgens stelt het CvTE jaarlijks vast of de erkende doorstroomtoets nog voldoet aan de criteria van het beoordelingskader. Dit doet het CvTE met de resultaten van de door een adviseur uitgevoerde jaarlijkse vaststelling. Onderdeel van de jaarlijkse vaststelling is de controle achteraf of de toetssamenstellingsprocedure door alle toetsaanbieders correct en consistent was uitgevoerd.

Checklist 1 Toetsopgaven Nederlandse taal per domein en per onderdeel: formulier voor toetsontwikkelaar en expert

A. Subdomein Zakelijke Teksten

Taken

Code	Teksten
L1.1	Informatieve teksten
L1.2	Instructieve teksten
L1.3	Betogende teksten

⁷ Zie Tabel 14 in: Glas, C.A.W., & Emons, W.H.M. (2017). Blueprint voor psychometrische verantwoording normering toetsadviezen en ijking op de referentieniveaus. EPO: Utrecht



Tekstkenmerken

Code		Mogelijke antwoorden	
L2.1	De tekstkenmerken van de teksten.		
Criteria waarop gescoord kan worden ¹		Onvoldoende	Voldoende
Inhoud	Er wordt niet teveel nieuwe informatie gelijktijdig geïntroduceerd.		
Structuur	De structuur van de teksten is eenvoudig en helder.		
Taalgebruik	De teksten hebben een lage informatiedichtheid.		
	De teksten bestaan voornamelijk uit frequent gebruikte woorden.		
	De teksten zijn niet te lang		
Bedoeling	De informatie in de teksten is herkenbaar geordend; verbanden in de tekst zijn duidelijk aangegeven.		

¹ Globaal oordeel gebaseerd op het totale aantal teksten in de toets. Deze criteria zijn bedoeld als hulpmiddel om tot een eendoordeel van de tekstkenmerken van alle in de toets opgenomen teksten te komen.

Kenmerken van de taakuitvoering

Code		Opgaven
L3.1	Opdrachten waarin kenmerk van de taakuitvoering: techniek en woordenschat wordt getoetst.	
L3.2	Opdrachten waarin kenmerk van de taakuitvoering: begrijpen wordt getoetst.	
L3.3	Opdrachten waarin kenmerk van de taakuitvoering: interpreteren wordt getoetst.	
L3.4	Opdrachten waarin kenmerk van de taakuitvoering: samenvatten wordt getoetst.	
L3.5	Opdrachten waarin kenmerk van de taakuitvoering: opzoeken wordt getoetst.	
L3.6	Opdrachten waarin kenmerk van de taakuitvoering: evalueren wordt getoetst (optioneel).	

B. Subdomein Fictionele, narratieve en literaire teksten

Taken

Code		Teksten
L4.1	Fictionele, narratieve en literaire teksten	

Tekstkenmerken

Code		Mogelijke antwoorden	
L5.1	De tekstkenmerken van de teksten.		
Criteria waarop gescoord kan worden ¹		Onvoldoende	Voldoende
Inhoud	Poëzie en liedjes hebben meestal een verhalende inhoud en een emotionele lading.		
Structuur	De structuur van de teksten is eenvoudig en helder.		
Taalgebruik	Het tempo waarin spannende of dramatische gebeurtenissen elkaar opvolgen is hoog.		
Bedoeling	Het verhaal heeft een dramatische verhaallijn waarin de spanning af en toe wordt onderbroken door gedachten of beschrijvingen.		

¹ Globaal oordeel gebaseerd op het totale aantal teksten in de toets. Deze criteria zijn bedoeld als hulpmiddel om tot een eendoordeel van de tekstkenmerken van alle in de toets opgenomen teksten te komen.

Kenmerken van de taakuitvoering

Code		Opgaven
L6.1	Opdrachten waarin kenmerk van de taakuitvoering: begrijpen wordt getoetst.	
L6.2	Opdrachten waarin kenmerk van de taakuitvoering: interpreteren wordt getoetst.	
L6.3	Opdrachten waarin kenmerk van de taakuitvoering: evalueren wordt getoetst (optioneel).	

B. Domein Taalverzorging

Kenmerken van de taakuitvoering:



Code	Opgaven
T1.1	Opdrachten waarin regels voor lettergreepgrenzen worden getoetst.
T1.2	Opdrachten waarin morfologische spelling wordt getoetst.
T1.3	Opdrachten waarin regels voor werkwoordspelling worden getoetst.
T1.4	Opdrachten waarin de overige regels worden getoetst.
T1.5	Opdrachten waarin het gebruik van leestekens wordt getoetst.
T1.6	Evenwichtige verdeling tussen opgaven rond niet-werkwoordspelling, werkwoordspelling en interpunctie.

Toetsopgaven

Evalueer de kwaliteit van alle toetsopgaven van de wettelijk verplichte domeinen Lezen en Taalverzorging (inclusief DICTEE) van het terrein Nederlandse taal, en de optionele domeinen Schrijven, Mondelinge taalvaardigheid, en Begrippenlijst van het terrein Nederlandse taal aan de hand van de kwaliteitseisen en constructievoorschriften uit de Checklist voor het beoordelen van de kwaliteit van observatie-categorieën en toetsopgaven.

Checklist 2 Toetsopgaven rekenen per domein en per onderdeel: formulier voor toetsontwikkelaar en expert

	Code	A Notatie, taal en betekenis	Code	B Met elkaar in verband brengen	Code	C Gebruiken
Getallen (g)	g.A.	Uitspraak, notatie en betekenis van gehele getallen, breuken en decimale getallen Uitspraak, notatie en betekenis van symbolen en relaties Uitspraak, taal en betekenis van wiskundetaal binnen het domein getallen	g.B.	Getallen en getalsrelaties Structuur en samenhang zien binnen het domein getallen	g.C.	Memoriseren, automatiseren Hoofdrekenen (notaties toegestaan) Hoofdbewerkingen (+,-,x,:) op papier uitvoeren met hele getallen en decimale getallen Bewerkingen met breuken (+,-,x,:) op papier uitvoeren Berekeningen uitvoeren om problemen op te lossen met gehele getallen, breuken en decimale getallen Rekenmachine op een verstandige manier inzetten

Toetsopgave

	Code	A Notatie, taal en betekenis	Code	B Met elkaar in verband brengen	Code	C Gebruiken
Verhoudingen (vh)	vh.A.	Uitspraak, notatie en betekenis van breuken, procenten en verhoudingen, Uitspraak, notatie en betekenis van symbolen en relaties Uitspraak, taal en betekenis van wiskundetaal binnen het domein verhoudingen	vh.B.	Verhouding, procent, breuk, decimaal getal, deling, 'deel van' met elkaar in verband brengen	vh.C.	In de context van verhoudingen berekeningen uitvoeren, ook met procenten en verhoudingen

Toetsopgave

	Code	A Notatie, taal en betekenis	Code	B Met elkaar in verband brengen	Code	C Gebruiken
Metten & Meetkunde (m/mk)	m/mk.A	Uitspraak, notatie en betekenis van maten voor lengte, oppervlakte, inhoud en gewicht, temperatuur, tijd en geld, Uitspraak, notatie en betekenis van meetkundige symbolen en relaties Meetinstrumenten aflezen	m/mk.B	Meetinstrumenten gebruiken Structuur en samenhang tussen maateenheden Verschillende meetkundige representaties	m/mk.C	Metten en rekenen met maten Rekenen in de meetkunde

Toetsopgave



	Code	A Notatie, taal en betekenis	Code	B Met elkaar in verband brengen	Code	C Gebruiken
Verbanden (vb)	vb.A.	Analyseren en interpreteren van informatie uit tabellen, grafische voorstellingen en beschrijvingen Kennis van diagrammen, tabellen en grafieken	vb.B.	Verschillende voorstellingsvormen met elkaar in verband brengen Gegevens verzamelen, ordenen en weergeven Patronen beschrijven	vb.C.	Tabellen, diagrammen en grafieken gebruiken bij het oplossen van problemen Rekenvaardigheden gebruiken binnen het domein verbanden

Toetsopgave

Checklist 3 Verdeling van toetsopgaven rekenen over domeinen en over onderdelen volgens de Toetswijzer PO

	A Notatie, taal en betekenis [geen minimum]	B Met elkaar in verband brengen [ten minste 20% van de gehele toets]	C Gebruiken [ten minste 20% van de gehele toets]
Getallen 30% – 40% van gehele toets	Uitspraak, notatie en betekenis van gehele getallen, breuken en decimale getallen, van symbolen en relaties Uitspraak, taal en betekenis van wiskundetaal binnen het domein getallen	Getallen en getsrelaties Structuur en samenhang	Memoriseren, automatiseren Hoofdrekenen (notaties toegestaan) Hoofdbewerkingen (+, -, x, :) op papier uitvoeren met hele getallen en decimale getallen Bewerkingen met breuken (+, -, x, :) op papier uitvoeren Berekeningen uitvoeren om problemen op te lossen met gehele getallen, breuken en decimale getallen Rekenmachine op een verstandige manier inzetten
Verhoudingen 20% – 30% van gehele toets	Uitspraak, notatie en betekenis van breuken, procenten en verhoudingen, Uitspraak, notatie en betekenis van symbolen en relaties Uitspraak, taal en betekenis van wiskundetaal binnen het domein verhoudingen	Verhouding, procent, breuk, decimaal getal, deling, 'deel van' met elkaar in verband brengen	In de context van verhoudingen berekeningen uitvoeren, ook met procenten en verhoudingen
Metten & Meetkunde 20% – 30% van gehele toets	Uitspraak, notatie en betekenis van maten voor lengte, oppervlakte, inhoud en gewicht, temperatuur, tijd en geld Uitspraak, notatie en betekenis van meetkundige symbolen en relaties Notatie, taal en betekenis van meetinstrumenten	Meetinstrumenten gebruiken Structuur en samenhang tussen maateenheden Verschillende meetkundige representaties	Metten en rekenen met maten Rekenen in de meetkunde
Verbanden 15% – 20% van totaal aantal toetsopgaven	Analyseren en interpreteren van informatie uit tabellen, grafische voorstellingen en beschrijvingen Kennis van diagrammen, tabellen en grafieken	Verschillende voorstellingsvormen met elkaar in verband brengen Gegevens verzamelen, ordenen en weergeven Patronen beschrijven	Tabellen, diagrammen en grafieken gebruiken bij het oplossen van problemen Rekenvaardigheden gebruiken binnen het domein verbanden



TOELICHTING

Algemeen

Sinds de invoering van de Wet centrale eindtoets en leerling- en onderwijsvolgsysteem primair onderwijs (Stb. 2014, 13) is het bevoegd gezag in het primair onderwijs verplicht om een centrale eindtoets in het achtste schooljaar af te nemen.

Op 1 januari 2023 treedt de Wet van 9 februari 2022 tot wijziging van een aantal onderwijswetten in verband met aanpassingen op het gebied van de doorstroom van het basisonderwijs naar het voortgezet onderwijs en wijziging van de stelselinrichting van doorstroomtoetsen en toetsen verbonden aan leerling- en onderwijsvolgsystemen in het basisonderwijs (Stb. 2022, 135) in werking. Met de inwerkingtreding van deze wet krijgt het College voor toetsen en examens de taak om de doorstroomtoets voor een periode van vier jaren te erkennen en steeds jaarlijks vast te stellen of de toets die wordt aangeboden voldoet aan de voorwaarden op basis waarvan de erkenning is verleend. Het College beoordeelt of de toetsen voldoen aan de in de Wet op het primair onderwijs, de Wet primair onderwijs BES, onderscheidenlijk de Wet op de expertisecentra, en het Toetsbesluit PO vastgelegde voorschriften.

Artikel 1

Op grond van artikel 3a, eerste lid, onderdeel g, van de Wet College voor toetsen en examens is het College vanaf 1 januari 2023 bevoegd om bij regeling een beoordelingskader vast te stellen voor erkenning van de doorstroomtoets en voor de jaarlijkse vaststelling of de erkende doorstroomtoets voldoet. Dit beoordelingskader wordt met deze regeling vastgelegd.

Het bevoegd gezag is verplicht om in ieder geval gebruik te maken van een door het College erkende doorstroomtoets voor de onderdelen Nederlandse taal en rekenen. Deze onderdelen moeten op grond van artikel 45b, vierde lid, van de Wet op het primair onderwijs, artikel 48c, vierde lid, van de Wet op de expertisecentra en 51b, tweede lid, van de Wet primair onderwijs BES verplicht getoetst worden. Het College kan ook de onderdelen van de toets beoordelen waarvoor niet is voorgeschreven dat zij onderdeel uitmaken van een doorstroomtoets.

Artikel 2

De gedetailleerde uitwerkingen voor de uitvoering van de leveringen van de gegevens, bedoeld in de paragrafen 5.2 en 5.3 van de bijlage, worden beschreven in een handboek doorstroomtoets en een handboek normering zoals gepubliceerd op de website van het College voor toetsen en examens (cvte.nl/onderwerpen/toetsen-primair-onderwijs).

Aanlevering van de gegevens op de wijze beschreven in deze handboeken stelt het College in staat het besluitvormingsproces zorgvuldig te doorlopen zonder dat dit proces moet worden onderbroken om de aanbieders in staat te stellen de gegevens aan te vullen. De beslistermijn voor het College is relatief kort en besluitvorming moet voor alle aanvragers tegelijkertijd worden uitgevoerd omdat de aanbieders tegelijkertijd hun aanbod op orde willen hebben. Ook de logistiek van het proces waarbij een commissie en een adviseur worden betrokken tijdens de zomerperiode vergt dat de gegevens tijdig, compleet en op vergelijkbare wijze worden aangeboden. Het is dus van belang voor de aanbieders om kennis te nemen van de procesafspraken en deze na te komen. Op de website van het College voor toetsen en examens (cvte.nl/onderwerpen/toetsen-primair-onderwijs) wordt duidelijk aangegeven welke handboeken moeten worden toegepast.

Artikel 3

Deze regeling treedt in werking tegelijkertijd met artikel VIII, onderdeel B, van de Wet van 9 februari 2022 tot wijziging van een aantal onderwijswetten in verband met aanpassingen op het gebied van de doorstroom van het basisonderwijs naar het voortgezet onderwijs en wijziging van de stelselinrichting van doorstroomtoetsen en toetsen verbonden aan leerling- en onderwijsvolgsystemen in het basisonderwijs (Stb. 2022, 135) in werking treedt. De beoogde inwerkingtredingsdatum is 1 januari 2023.

*Het College voor toetsen en examens, de voorzitter,
J.H. van der Vegt*